

\*\*\*\*\*

(TM)

\*\*\*\*\*

Release 3.1A John F. Collins, Biocomputing Research Unit.  
Copyright (c) 1993-1998 University of Edinburgh, U.K.  
Distribution rights by Oxford Molecular Ltd

MPsrch\_nm n.a. - n.a. database search, using Smith-Waterman algorithm  
Run on: Thu Apr 20 16:14:23 2000; MasPar time 6236.81 Seconds  
Tabular output not generated. 1390.567 Million cell updates/sec

Title: >US-09-249-003-1  
Description: (1-2297) from US09249003.seq  
Perfect Score: 2297  
N.A. Sequence: 1 CGCGTGTGTTGCCGACGCTT.....TAATCCATGTTCTTTTCC 2297  
Comp: GCCGACACACGGCTCAGAA.....ATTAGGTACAAAGAAAGG

Scoring table:

Gap 6

Nmatch STD : Dbase 0; Query 0

Searched: 4538634 seqs, 1887831982 bases x 2

Post-processing: Minimum Match 0%  
Listing first 45 summaries

Database:

embl-est61  
1:em\_est15 2:em\_est23 3:em\_est29 4:em\_est3 5:em\_est30  
6:em\_est31 7:em\_est33 8:em\_est34 9:em\_est5 10:em\_gss1  
11:em\_gss12

Database:

genbank-est115  
12:gb\_est1 13:gb\_est10 14:gb\_est11 15:gb\_est12  
16:gb\_est13 17:gb\_est14 18:gb\_est15 19:gb\_est16  
20:gb\_est17 21:gb\_est18 22:gb\_est19 23:gb\_est2  
24:gb\_est20 25:gb\_est21 26:gb\_est22 27:gb\_est23  
28:gb\_est24 29:gb\_est25 30:gb\_est26 31:gb\_est27  
32:gb\_est28 33:gb\_est29 34:gb\_est3 35:gb\_est30  
36:gb\_est31 37:gb\_est32 38:gb\_est33 39:gb\_est34  
40:gb\_est35 41:gb\_est36 42:gb\_est37 43:gb\_est38  
44:gb\_est39 45:gb\_est4 46:gb\_est40 47:gb\_est41  
48:gb\_est42 49:gb\_est43 50:gb\_est44 51:gb\_est5  
52:gb\_est6 53:gb\_est7 54:gb\_est8 55:gb\_est9 56:gb\_gss1  
57:gb\_gss10 58:gb\_gss11 59:gb\_gss12 60:gb\_gss13  
61:gb\_gss14 62:gb\_gss15 63:gb\_gss2 64:gb\_gss3 65:gb\_gss4  
66:gb\_gss5 67:gb\_gss6 68:gb\_gss7 69:gb\_gss8 70:gb\_gss9

Statistics: Mean 12.223; Variance 2.355; scale 5.190

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	ID	Description	Pred. No.
1	736	32.0	750	26	AT042325 oy37e10.x1 Soares_para	0.00e+00
2	721	31.4	752	27	A1089762 qa22e08.x1 NCI_CGAP.Br	0.00e+00
3	713	31.0	777	21	AA701678 z143h11.s1 Soares_feta	0.00e+00

## ALIGNMENTS

RESULT	1	AI042325	750 bp	mrna	EST	30-JUN-1998
LOCUS	oy37e10.x1 Soares_parathyroid_tumor_NHHPA Homo sapiens cDNA clone					
DEFINITION	IMAGE:1668042 3' similar to gb:M58342 IDURONATE 2-SULFATASE					
PRECUSOR	(HUMAN);, mRNA sequence.					
ACCESSION	AI042325					
VERSION	AI042325.1					
KEYWORDS	EST.					
SOURCE	human.					
ORGANISM	Homo sapiens					
REFERENCE	1 (bases 1 to 750)					
AUTHORS	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;					
TITLE	Eutheria; Primates; Catarrhini; Hominidae; Homo.					
JOURNAL	NCI-CGAP <a href="http://www.ncbi.nlm.nih.gov/ncicgap">http://www.ncbi.nlm.nih.gov/ncicgap</a> .					
COMMENT	National Cancer Institute, Cancer Genome Anatomy Project (CGAP), Tumor Gene Index					
	Unpublished (1997)					
	On Jan 19, 1998 this sequence version replaced gi:2151624.					
	Contact: Robert Strausberg, Ph.D.					
	Tel: (301) 496-1550					
	Email: Robert_Strausberg@nih.gov					
	cDNA Library Preparation: M. Bento Soares, Ph.D., M. Fatima					
	Bonaldo, Ph.D.					
	cDNA Library Arrayed by: Greg Lennon, Ph.D.					
	DNA Sequencing by: Washington University Genome Sequencing Center					
	Clone distribution: NCI-CGAP clone distribution information can be					
	found through the I.M.A.G.E. Consortium/LLNL at:					

786 40	AI819115	wj79c02.x1	NCI_CGAP_Lu	0.00e+00
786 33	AI553770	tn28g04.x1	NCI_CGAP.Br	0.00e+00
694 37	AI744724	wf90d04.x1	NCI_CGAP.Co	0.00e+00
789 22	AA805711	ns42c05.s1	NCI_CGAP_GC	0.00e+00
653 50	AW173232	x185a05.x1	Soares_NFL	0.00e+00
704 19	AA553786	n135d03.s1	NCI_CGAP_Lu	0.00e+00
606 18	AA513794	nh89c05.r1	NCI_CGAP.Br	0.00e+00
644 13	AA131455	z132d04.r1	Soares_preg	0.00e+00
629 32	AI478887	tn25b02.x1	Soares_NFL	0.00e+00
576 14	AA236900	z843h10.s1	Soares_NHML	0.00e+00
566 38	AI769955	wj22f11.x1	NCI_CGAP_Ki	0.00e+00
551 31	AI418603	tg37d08.r1	Soares_NFL	0.00e+00
556 17	AA399142	z149f03.r1	Soares_ovar	0.00e+00
604 26	AI033665	ov22g07.x1	Soares_para	0.00e+00
651 20	AA642413	ns22g07.s1	NCI_CGAP_GC	0.00e+00
552 18	AA506162	n114a10.s1	NCI_CGAP.Co	0.00e+00
552 32	AI471587	ta15f06.x1	NCI_CGAP_Ly	0.00e+00
496 27	AI143191	qb72a12.x1	Soares_feta	0.00e+00
712 39	AI798738	we92a07.x1	Soares_NFL	0.00e+00
986 32	AI524108	th09c01.x1	NCI_CGAP_CL	0.00e+00
526 48	AW160323	au66a10.y1	Schneider_f	0.00e+00
570 29	AI271813	qj88a08.x1	NCI_CGAP_Ki	0.00e+00
589 31	AI400170	tg67h06.x1	Soares_NHML	0.00e+00
504 25	AA935580	on13d04.s1	NCI_CGAP_Lu	0.00e+00
635 34	H06357	y179b02.x1	Soares_infa	0.00e+00
469 30	AI359729	qy28f03.x1	NCI_CGAP.Br	0.00e+00
511 28	AI186134	q634h05.s1	Soares_feta	0.00e+00
516 29	AI283859	qk46c03.x1	NCI_CGAP.Co	0.00e+00
466 28	AI174207	on01b09.x1	Stratagene	0.00e+00
480 29	AI292239	qm76a02.x1	Soares_plac	0.00e+00
453 27	AI096538	q661d01.x1	NCI_CGAP.Br	0.00e+00
443 30	AI359745	qy28h03.x1	NCI_CGAP.Br	0.00e+00
455 44	AW134717	UI-H-Bil-abq-a-11-0-UI		0.00e+00
437 30	AI359737	qy28g03.x1	NCI_CGAP.Br	0.00e+00
519 21	AA700596	z141b07.s1	Soares_feta	0.00e+00
479 51	N26477	yy61b05.s1	Soares_mela	0.00e+00
452 25	AA993642	ot84g10.s1	Soares_tota	0.00e+00
460 13	AA131408	z132d04.s1	Soares_preg	0.00e+00
467 36	AI678627	tu84h05.x1	NCI_CGAP_Ga	0.00e+00
517 34	H08043	y191g05.r1	Soares_infa	0.00e+00

Query Match 31.4%; Score 721; DB 27; Length 752;  
Best Local Similarity 98.5%; Pred. No: 0.00e+00;  
Matches 741: Conservative 0; Mismatches 9; Indels

Query Match 31.4%; Score 721; DB 27; Length 752;  
Best Local Similarity 98.5%; Pred. No. 0.00e+00;  
Matches 741: Conservative 0; Mismatches 9; Indels 2; Gaps 2.

Db	1	AAATGCGGCCACCCGGACCGCGGAGGCC	TTCTCTGGCTGGGTCTGTTCTGAGCTCGG	60
Qy				
Db	123	AAATGCGGCCACCCGGACCGCGGAGGCC	TTCTCTGGCTGGGTCTGTTCTGAGCTCGG	182
Qy				
Db	61	TCTGGGTGGCCCTGGGATCCGAAAGCGC	AGGCCAACTCGACCACACAGATGCTCTGA	120
Qy				
Db	183	TCTGGGTGGCCCTGGGATCCGAAAGCGC	AGGCCAACTCGACCACACAGATGCTCTGA	242
Qy				
Db	121	TTCTCATCATCGTGGATGACCTGGCCG	CTCCCTGGGCTGTTATGGGATAGCTGGTGA	180
Qy				
Db	243	TTCTCATCATCGTGGATGACCTGGCCG	CTCCCTGGGCTGTTATGGGATAGCTGGTGA	302
Qy				
Db	181	GGTCCCAAAATATTGACCAACTGGCAT	CCCCACAGGCTCTCTTCAGAAATGCTTTGG	240
Qy				
Db	303	GGTCCCAAAATATTGACCAACTGGCAT	CCCCACAGGCTCTCTTCAGAAATGCTTTGG	362
Qy				
Db	241	AGCAAGCAGTGTGGCCCCCGAGCGCGT	TTCTTTCTCCTCAGTGGCAGGAGACCTGAC	300
Qy				
Db	363	AGCAAGCAGTGTGGCCCCCGAGCGCGT	TTCTTTCTCCTCAGTGGCAGGAGACCTGAC	422
Qy				
Db	301	CCCGCCTGTACGACTTCAACTTCAACT	TCTTACTGGAGGGTGCAGCTTGGAAACTT	360
Qy				
Db	423	CCCGCCTGTACGACTTCAACTTCAACT	TCTTACTGGAGGGTGCAGCTTGGAAACTT	482
Qy				
Db	361	CCCAAGTACTTCAAGGAGAAATGGCTAT	GTGACATGCTGGTGGGAAAAGTCTTTTCA	420
Qy				
Db	483	CCCAAGTACTTCAAGGAGAAATGGCTAT	GTGACATGCTGGTGGGAAAAGTCTTTTCA	542
Qy				
Db	421	GGATATCTTCTAAACATACCGATGATT	CTCCGTATAGCTGGTCTTTTCCACCTTAT	480
Qy				
Db	543	GGATATCTTCTAAACATACCGATGATT	CTCCGTATAGCTGGTCTTTTCCACCTTAT	602
Qy				
Db	481	CTTCTCTGAGAAGTATGAAAACACTA	AGACATGTCTGAGGGCCAGATGGAGAACT	540
Qy				
Db	603	CTTCTCTGAGAAGTATGAAAACACTA	AGACATGTCTGAGGGCCAGATGGAGAACT	662
Qy				
Db	541	CCAACTTGCCTTGGCCTGTGGATGTCT	GGATGTTCCCGAGGGCACCTTTGCTGACA	600
Qy				
Db	663	CCAACTTGCCTTGGCCTGTGGATGTCT	GGATGTTCCCGAGGGCACCTTTGCTGACA	722
Qy				
Db	601	AGACACTTGNAGCAGCCATACAGTTGT	TGGAAAAGATGAAAACGTCAGCAGTCTCT	660
Qy				
Db	723	AGACACTTGNAGCAGCCATACAGTTGT	TGGAAAAGATGAAAACGTCAGCAGTCTCT	782
Qy				
Db	661	TCCTTGGCGGTTGGGTATCATTAAGCC	ACACCATCCCTTTCAGATACCCCGAGGA	720
Qy				
Db	783	TCCTTGGCGGTTGGGTATCATTAAGCC	ACACCATCCCTTTCAGATACCCCGAGGA	841
Qy				
Db	721	AAAGTGTATCCCTTTGGAGAACATCAC	CCCTGG	752
Qy				
Db	842	AAAGTGTATCCCTTTGGAGAACATCAC	CCCTGG	872
Qy				

RESULT	3	AA701678	777 bp	mrna	EST	19-DEC-1997
LOCUS		z143h11.s1	Soares fetal_liver_spleen	INFLS.S1	Homo sapiens	CDNA
DEFINITION		IMAGE:1433605	3' similar to gb:M58342	IDURONATE 2-SULFATASE		
		PRECURSOR (HUMAN);	mrna sequence.			
ACCESSION		AA701678				
VERSION		AA701678.1	GI:2704843			
KEYWORDS		EST.				
SOURCE		human.				
ORGANISM		Homo sapiens				
		Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;				
		Eutheria; Primates; Catarrhini; Hominidae; Homo.				
REFERENCE		1 (bases 1 to 777)				
AUTHORS		Hillier L., Allen M., Bowles L., Dubuque T., Geisel G., Jost S.,				
		Krizman D., Kucaba T., Lacy M., Le N., Lennon G., Marra M.,				
		Martin J., Moore B., Schellenberg K., Steptoe M., Tan F.,				
		Theising J., White Y., Willie T., Waterston R. and Wilson R.				
TITLE		WashU-NCI human EST Project				
JOURNAL		Unpublished (1997)				

On Sep 12, 1996 this sequence version replaced gi:1393828.  
Contact: Wilson RK  
Washington University School of Medicine  
4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108  
tel: 314 286 1800  
Fax: 314 286 1810  
Email: estewatson.wustl.edu  
This clone is available royalty-free through LNL ; contact the  
IMAGE Consortium (info@image.llnl.gov) for further information.  
Seq primer: -40ml3 fwd. ET from Amersham  
High quality sequence stop: 452.

## FEATURES

Source	1..777	234 a	175 c	179 g	189 t
source	/db_xref="GDB:133374" /db_xref="taxon:9606" /clone="IMAGE:433605" /clone_lib="Soares_fetal_liver_spleen_INFLS_S1" /sex="male" /dev_stage="20 week-post conception fetus" /lab_host="DH10B (ampicillin resistant)" /note="Organ: Liver and Spleen; Vector: pT7T3D (Pharmacia) with a modified polylinker; Site.1: Pac I; Site.2: Eco RI; This is a subcloned version of the original Soares fetal liver spleen INFLS library. 1st strand cDNA was primed with a Pac I - oligo(dT) primer [5', AACTGGAGAAGTAATTAAGAGCTTTTTTTTTTTTTTTT 3'], double-stranded cDNA was ligated to Eco RI adaptors (Pharmacia), digested with Pac I and cloned into the Pac I and Eco RI sites of the modified pT7T3 vector. Library went through one round of normalization. Library constructed by Bento Soares and M.Fatima Bonaldo." BASE COUNT 234 a 175 c 179 g 189 t ORIGIN				
Query Match	31.0%; Score 713; DB 21; Length 777;				
Best Local Similarity	97.2%; Pred. No. 0.00e+00;				
Matches	755; Conservative 0; Mismatches 18; Indels 4; Gaps 4;				
Db	1	ACCAATTATAATTTTAATAAGACTAAACGAAAGGTTTGGCTGTTCACATATTCTCAGG	60		
Cp	1980	ACCAATTATAATTTTAATAAGACTAAACGAAAGGTTTGGCTGTTCACATATTCTCAGG	1921		
Db	61	CCAAATTGTCATGCTGTTGGATTAACTAGCCCTCAGCTGCTTCCAAATATTATGGGT	120		
Cp	1920	CCAAATTGTCATGCTGTTGGATTAACTAGCCCTCAGCTGCTTCCAAATATTATGGGT	1861		
Db	121	AATCACAAACGACACAGCTCTTAACCTCCTCTCTCACACAGCTGGAAGGAGCACATCAT	180		
Cp	1860	AATCACAAACGACACAGCTCTTAACCTCCTCTCTCACACAGCTGGAAGGAGCACATCAT	1801		
Db	181	TGGCATCCATGTTGGCAAACTCAAGGCATCAACAACTGGAAAGATCTCCACCTTGG	240		
Cp	1800	TGGCATCCATGTTGGCAAACTCAAGGCATCAACAACTGGAAAGATCTCCACCTTGG	1741		
Db	241	GAATCATTTACATATTCTGTATCTGCAATGGGTCAGAAATCCACAAAATACAGTCCCT	300		
Cp	1740	GAATCATTTACATATTCTGTATCTGCAATGGGTCAGAAATCCACAAAATACAGTCCCT	1681		
Db	301	GCATGGATGTCAGAAAAGTTAGCTAGAAATTCATCAGGATTGAAGCCCAACCCACACAGTA	360		
Cp	1680	GCATGGATGTCAGAAAAGTTAGCTAGAAATTCATCAGGATTGAAGCCCAACCCACACAGTA	1621		
Db	361	TACCTATAGTCTATGGTGGCGTATGGAATAGCCCATGATCTTTTATATCTTTTAAACTCGGC	420		
Cp	1620	TACCTATAGTCTATGGTGGCGTATGGAATAGCCCATGATCTTTTATATCTTTTAAACTCGGC	1561		
Db	421	TTGTCAGAAATTCACATGAGGATGCTGAAAGCCGGGATACCTGGCTATAGGCAATCAGT	480		
Cp	1560	TTGTCAGAAATTCACATGAGGATGCTGAAAGCCGGGATACCTGGCTATAGGCAATCAGT	1501		
Db	481	TCACGGGATTTCACGAGGAGGTACGGATCTCTTCCAAAGTCACGGAATCGAAATTCGCTTC	540		

```

Cp 1500 TCACGGGATTACACGGAGGTACGGATCCTCTTCCAAAGTCACGGAATCGAAATGCTTC 1441
Db 541 AGAAGGTTCTTGCTTCTGTCACAGCTCAACGTGAATGAAGAACGGGCGCAGGTT 600
Cp 1440 AGAAGGTTCTTGCTTCTGTCACAGCTCAACGTGAATGAAGAACGGGCGCAGGTT 1381
Db 601 GGAACCTGCAGTCTGCAAGTTACAGCAGCGTTGGGAAAGAGACAAAGTTCACACA 660
Cp 1380 GGAACCTGCAGTCTGCAAGTTCCAGCGCAGCGT-GGGAAAGAGACAAAGTTCACACA-A 1323
Db 661 GGTTCATGGATTGGCCCTGCTGCCATTAACCTTGAGGCGCAATCAAAAGGGTCGAGG 720
Cp 1322 GTCCATGGATTGGCTGCTGCCCTGCC-ATCAACTGTGAGGCGGAATCAAAAGGGTCGAGG 1264
Db 721 TAAGGGAAAGCTTCTCGCTGTTCCGGAAGTGAAGACGTTCTTCCAAAGAAACATA 777
Cp 1263 TAAGGGAAAGCTTCTCGCTGCTCCGGAAGTGAAGCGCTCCTTCCA-GGAACATA 1208

RESULT 4
LOCUS AI819115 786 bp mRNA EST 24-AUG-1999
DEFINITION wJ79G02.x1 NCI_CGAP_Lu19 Homo sapiens cDNA clone IMAGE:2409026 3',
similar to gb:M58342 IDURONATE 2-SULFATASE PRECURSOR (HUMAN);, mRNA
sequence.
ACCESSION AI819115
VERSION AI819115.1 GI:5438118
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 786)
AUTHORS NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.
TITLE National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
Tumor Gene Index
JOURNAL Unpublished (1997)
COMMENT On Jun 5, 1998 this sequence version replaced gi:3189120.
Contact: Robert Strausberg, Ph.D.
Tel: (301) 496-1550
Email: Robert.Strausberg@nih.gov
Tissue Procurement: Christopher Moskaluk, M.D., Ph.D., Michael R.
Emmert-Buck, M.D., Ph.D.
cDNA Library Preparation: M. Bento Soares, Ph.D.
DNA Library Arrayed by: Greg Lennon, Ph.D.
Sequencing by: Washington University Genome Sequencing Center
Clone distribution: NCI-CGAP clone distribution information can be
found through the I.M.A.G.E. Consortium/LINL at:
www-bio.llnl.gov/dbp/image/image.html

Seq primer: -400P from Gibco
High quality sequence stop: 434.
FEATURES
Location/Qualifiers
source 1..786
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:2409026"
/clone_lib="NCI_CGAP_Lu19"
/tissue_type="squamous cell carcinoma, poorly
differentiated (4 pooled tumors, including primary and
metastatic)"
/dev_stage="adult"
/lab_host="DH10B (phage-resistant)"
/notes="Organ: lung; Vector: pT73D-Pac (Pharmacia) with a
modified polylinker; 1st strand cDNA was prepared from
pooled lung tumor tissue, and was then primed with a Not I
- oligo(dT) primer. Double-stranded cDNA was ligated to
Eco RI adaptors (Pharmacia), digested with Not I and
cloned into the Not I and Eco RI sites of the modified
pT73 vector. Library went through one round of
normalization. Library constructed by Bento Soares and M.
Fatima Bonaldo."
BASE COUNT 175 a 242 c 182 g 179 t 8 others
ORIGIN
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;

```

```

Query Match 30.8%; Score 708; DB 40; Length 786;
Best Local Similarity 96.8%; Pred. No. 0.00e+00;
Matches 766; Conservative 0; Mismatches 17; Indels 8; Gaps 8;

Db 1 GCTGGTCTGTGTTCTGAGCTCGTCTGCGCTCGCCCTCGGATCCGAAACCGACGCCAATC 60
QY 160 GCTGGTCTGTGTTCTGAGCTCGTCTGCGCTCGCCCTCGGATCCGAAACCGACGCCAATC 219
Db 61 GACCACAGATGCTCTGAAGTTCTTCTCATCATCGTGGATGACCTGCGCCCTCCCTGGG 120
QY 220 GACCACAGATGCTCTGAAGTTCTTCTCATCATCGTGGATGACCTGCGCCCTCCCTGGG 279
Db 121 CTGTTATGGGATAAGCTGGTGGTGTCGCCAATAATTTGACCAACTGGCATCCACAGCCT 180
QY 280 CTGTTATGGGATAAGCTGGTGGTGTCGCCAATAATTTGACCAACTGGCATCCACAGCCT 339
Db 181 CCTCTTCCAGATGCTTTGGCGCAGCAGTGTGCGCCCGGAGCGGCTTCTTTCTCT 240
QY 340 CCTCTTCCAGATGCTTTGGCGCAGCAGTGTGCGCCCGGAGCGGCTTCTTTCTCT 399
Db 241 CACTGGCAGGACCTGACACACACCGCCTGTACGACTTCAACTCCTACTGGAGGTGCA 300
QY 400 CACTGGCAGGACCTGACACACACCGCCTGTACGACTTCAACTCCTACTGGAGGTGCA 459
Db 301 CGCTGGAACCTTCTCCACCATCCCGCAGTACTCAAGGAGAATGGCTATGTGACCATGTC 360
QY 460 CGCTGGAACCTTCTCCACCATCCCGCAGTACTCAAGGAGAATGGCTATGTGACCATGTC 519
Db 361 GGTGGGAAAAGTCTTTACCCCTGGGATATCTTCTTAACCATACCGATGATTCCTCGTATAG 420
QY 520 GGTGGGAAAAGTCTTTACCCCTGGGATATCTTCTTAACCATACCGATGATTCCTCGTATAG 579
Db 421 CTGCTCTTTTCCACCTTATCATCTTCTCTGAGAGTATGAAACACATAAGACATGTCG 480
QY 580 CTGCTCTTTTCCACCTTATCATCTTCTCTGAGAGTATGAAACACATAAGACATGTCG 639
Db 481 AGGGCCAGATGGAGAATCCATGCCCAACCTGCTNTGCCCTGGTGGATGTGCTG-ATGTTC 539
QY 640 AGGGCCAGATGGAGAATCCATGCCCAACCTGCTNTGCCCTGGTGGATGTGCTG-ATGTTC 698
Db 540 CGAGGGCACTTGGCTGCACAAACAGACAGCACTGAGNAGCCCATACAGTTGTTGGAAAGA 599
QY 699 CGAGGGCACTTGGCTGCACAAACAGACAGCACTGAGNAGCCCATACAGTTGTTGGAAAGA 758
Db 600 TGAAAACGTACCCAGTCTTTCTTCNCTGCGCGGGTATCATAGCCACACATNCCCT 659
QY 759 TGAAAACGTACCCAGTCTTTCTTCTCTGCGCGTTCATAGAGCCACATCCCT 818
Db 660 TCAGATACCCCAAGAAATTTCCAGAAGT-GTATCCCTGGAGAACATCACCNTGGCCCC 718
QY 819 TCAGATACCCCAAGAAATTTCCAGAAGTGTATCCCTTGGAGAACATCACC-CTGGCCCC 876
Db 719 CGATCCCGAG-TNCTGTGATGGGTACCCCTGTG-CCTACAACCCCTGGATG-ACATCAN 775
QY 877 CGATCCCGAGTCCCTGTGATGGGTACCCCTGTG-CCTACAACCCCTGGATGATGACATCAG 936
Db 776 GCAACGGNAAG 786
QY 937 GCAACGGNAAG 947

RESULT 5
LOCUS AI553770 786 bp mRNA EST 12-MAY-1999
DEFINITION tn28g04.x1 NCI_CGAP_Brn25 Homo sapiens cDNA clone IMAGE:2168982 3',
similar to gb:M58342 IDURONATE 2-SULFATASE PRECURSOR (HUMAN);, mRNA
sequence.
ACCESSION AI553770
VERSION AI553770.1 GI:4486133
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;

```



REFERENCE  
AUTHORS  
TITLE  
JOURNAL  
COMMENT

Eutheria; Primates; Catarrhini; Homiidae; Homo.  
1 (bases 1 to 786)  
NCI/NINDS-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.  
National Cancer Institute / National Institute of Neurological  
Disorders and Stroke, Brain Tumor Genome Anatomy Project  
(CGAP/BTGP), Tumor Gene Index  
Unpublished (1997)  
On Mar 20, 1998 this sequence version replaced gi:2980487.  
Contact: Robert Strausberg, Ph.D.  
Tel: (301) 496-1550  
Email: Robert\_Strausberg@nih.gov  
Tissue Procurement: David N. Louis, M.D., Myrna R. Rosenfeld M.D.,  
Ph.D.  
cDNA Library Preparation: M. Bento Soares, Ph.D., M. Fatima  
Bonaldo, Ph.D.  
cDNA Library Arrayed by: Greg Lennon, Ph.D.  
DNA Sequencing by: Washington University Genome Sequencing Center  
Clone distribution: NCI-CGAP clone distribution information can be  
found through the I.M.A.G.E. Consortium/LLNL at:  
www-bio.llnl.gov/bbrp/image/image.html

Insert Length: 1027 Std Error: 0.00  
Seq primer: -400P from Gibco  
High quality sequence stop: 466  
POLYA=No.

FEATURES  
source  
Location/Qualifiers  
1..786  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/clone="IMAGE:2168982"  
/clone\_lib="NCI\_CGAP\_Brn25"  
/tissue\_type="anaplastic oligodendroglioma"  
/lab\_host="DH10B"  
/note="Organ: brain; Vector: pT7T3D-Pac (Pharmacia) with a  
modified polylinker; Site\_1: Not I; Site\_2: Eco RI; 1st  
strand cDNA was primed with a Not I - oligo(GT) primer [5'  
TGTTACCAATCTGAGTGGAGCGCGCGCATAGGTTTGTGTTTGTGTTT  
T 3']; double-stranded cDNA was ligated to Eco RI  
adaptors (Pharmacia), digested with Not I and cloned into  
the Not I and Eco RI sites of the modified pT7T3 vector.  
Library is normalized, and was constructed by Bento  
Soares and M. Fatima Bonaldo."

BASE COUNT 173 a 242 c 188 g 182 t 1 others  
ORIGIN

Query Match 30.4%; Score 699; DB 33; Length 786;  
Best Local Similarity 97.1%; Pred. No. 0.00e+00;  
Matches 761; Conservative 0; Mismatches 15; Indels 8; Gaps 8;

Db 1 GCGGCGGCTCGAAGCGGAAATGCGGCCACCGCGGACTTTCGAGGCGCTCTCTGCTGG 60  
Qy 105 GCGGCGGCTCGAAGCGGAAATGCGGCCACCGCGGAGCGGCGCTCTCTGCTGG 164  
Db 61 GCTCGGTTCTGAGCTCCGCTCTGCTGGCGCTGGATCCGAAGCGGCGGCACTCGACCA 120  
Qy 165 GCTCGGTTCTGAGCTCCGCTCTGCTGGCGCTGGATCCGAAGCGGCGGCACTCGACCA 224  
Db 121 CAGATGCTCTGAACGCTTCTTCATCATCGTGGATGACCTGGCGCCCTCCCTGGGCTGTT 180  
Qy 225 CAGATGCTCTGAACGCTTCTTCATCATCGTGGATGACCTGGCGCCCTCCCTGGGCTGTT 284  
Db 181 ATGGGGATAAGCTGGTGGAGTCCCAAAATATTGACCACTGGCATCCCAAGCGCTCTCT 240  
Qy 285 ATGGGGATAAGCTGGTGGAGTCCCAAAATATTGACCACTGGCATCCCAAGCGCTCTCT 344  
Db 241 TCCGAATGCCCTTTCGGCAGCAGAGTGTGGCGCCCGGAGCGCGTTCCTTCTCTCACTG 300  
Qy 345 TCCGAATGCCCTTTCGGCAGCAGAGTGTGGCGCCCGGAGCGCGTTCCTTCTCTCACTG 404  
Db 301 CGAGGAGACCTGACACACCGCCCTGTACGACTTCAACTCCCTACTGAGGCGTGCAGCGTG 360  
Qy 405 CGAGGAGACCTGACACACCGCCCTGTACGACTTCAACTCCCTACTGAGGCGTGCAGCGTG 464

Db 361 GAAACTTCTCCACCATCCCCAGTACTTCAAGAGAGATGGCTATGTGACCATGTGGTGG 420  
Qy 465 GAAACTTCTCCACCATCCCCAGTACTTCAAGAGAGATGGCTATGTGACCATGTGGTGG 524  
Db 421 GAAAGTCTTTACCCCTGGGATATCTTCTACCATACCGATGATTCCTCCGTATAGCTGGT 480  
Qy 525 GAAAGTCTTTACCCCTGGGATATCTTCTACCATACCGATGATTCCTCCGTATAGCTGGT 584  
Db 481 CTTTTCACCTTATCATCTTCTTCTGAGAGATGATAAAGACATTAAGACATGTCCAGGGC 540  
Qy 585 CTTTTCACCTTATCATCTTCTTCTGAGAGATGATAAAGACATTAAGACATGTCCAGGGC 644  
Db 541 CAGATGAGAACTCCATGCCAACCTGCTTTCCTGCTGGATGCTGGGATGTTCCGAGG 600  
Qy 645 CAGATGAGAACTCCATGCCAACCTGCTTTCCTGCTGGATGCTGGGATGTTCCGAGG 704  
Db 601 GCACCCCTGCTGACAA-CAGAGCAGCTGAGCAAGCCATACAGTTGTTGGAAAGATGAAA 659  
Qy 705 GCACCT-TGGCTGACAAACAGAGCAGCTGAGCAAGCCATACAGTTGTTGGAAAGATGAAA 763  
Db 660 ACCTCAAGCAGTCTTCTTCTTCTGCGCGT-GGGTATCATAGC-ACACATCCCTTACA 717  
Qy 764 ACCTCAAGCAGTCTTCTTCTGCGCGTGGGATCATAGCCACACATCCCTTACA 823  
Db 718 TACCCCAAGGAATT-CAGAGCTTG-ATCCTTTG-AGAACAT-ACCCTGGCGCCGATCCC 773  
Qy 824 TACCCCAAGGAATTTCAGAGTGTATCTCTTGGAGAACATCACCTTGGCGCCGATCCC 883  
Db 774 AAG 777  
Qy 884 GAG 887

RESULT 6  
LOCUS AI744724 694 bp mRNA EST 21-JUN-1999  
DEFINITION wf90404.x1 NCI-CGAP Co3 Homo sapiens cDNA clone IMAGE:2362855 3'  
similar to gb:M58342 IDURONATE 2-SULFATASE PRECURSOR (HUMAN);, mRNA  
sequence.  
ACCESSION AI744724  
VERSION AI744724.1 GI:5112934  
KEYWORDS EST.  
SOURCE human.  
ORGANISM Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;  
Eutheria; Primates; Catarrhini; Homiidae; Homo.  
REFERENCE 1 (bases 1 to 694)  
AUTHORS NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.  
TITLE National Cancer Institute, Cancer Genome Anatomy Project (CGAP),  
Tumor Gene Index  
JOURNAL Unpublished (1997)  
COMMENT On Dec 20, 1995 this sequence version replaced gi:1135683.  
Contact: Robert Strausberg, Ph.D.  
Tel: (301) 496-1550  
Email: Robert\_Strausberg@nih.gov  
Tissue Procurement: Elias Campo, M.D., Michael R. Emmert-Buck,  
M.D., Ph.D.  
cDNA Library Preparation: M. Bento Soares, Ph.D.  
cDNA Library Arraying: Greg Lennon, Ph.D.  
DNA Sequencing by: Washington University Genome Sequencing Center  
Clone distribution: NCI-CGAP clone distribution information can be  
found through the I.M.A.G.E. Consortium/LLNL at:  
www-bio.llnl.gov/bbrp/image/image.html

Seq primer: -400P from Gibco  
High quality sequence stop: 462.  
Location/Qualifiers  
1..694  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/clone="IMAGE:2362855"  
/clone\_lib="NCI\_CGAP\_Co3"  
/sex="pooled"  
/tissue\_type="colon"

FEATURES  
source  
Location/Qualifiers  
1..694  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/clone="IMAGE:2362855"  
/clone\_lib="NCI\_CGAP\_Co3"  
/sex="pooled"  
/tissue\_type="colon"



Db 322 AAAAAGTGGTCAATTAACCAATTATAAATTTTAATAAGACTTAACGAAAGGTTTGGCT 381  
|||||  
Cp 1996 AAAAAGTGGTCAATTAACCAATTATAAATTTTAATAAGACTTAACGAAAGGTTTGGCT 1937  
|||||  
Db 382 GTTACATATTTCTAGCGCAAAATGTTGATGATGTTGGATTAATAGCCCTCAGGCTGC 441  
|||||  
Cp 1936 GTTACATATTTCTAGCGCAAAATGTTGATGATGTTGGATTAATAGCCCTCAGGCTGC 1877  
|||||  
Db 442 TTCCAATATTATGGTATATCACAACACACAGCTCTAATCTCTCTCACCAGCTGA 501  
|||||  
Cp 1876 TTCCAATATTATGGTATATCACAACACACAGCTCTAATCTCTCTCACCAGCTGA 1817  
|||||  
Db 502 AGGAGCACATCACATTTGCCATCATCGTTGGCAAACTCAAGGCATCAACAACCTGGAA 561  
|||||  
Cp 1816 AGGAGCACATCACATTTGCCATCATCGTTGGCAAACTCAAGGCATCAACAACCTGGAA 1757  
|||||  
Db 562 AAGATCTCCCTTGGGAATCATTATACATATGTGATCCTGCAATGGGTCAAGATCCCC 621  
|||||  
Cp 1756 AAGATCTCCCTTGGGAATCATTATACATATGTGATCCTGCAATGGGTCAAGATCCAC 1697  
|||||  
Db 622 AAA-TAGGTTCCCTGACATGTTCCGAAAGTAGCTGGAATCC-TCCGGATTGA 679  
|||||  
Cp 1696 AAAATACAGTTCCCTGCA-TGGATGTGAGAAAGTTAGCTAGAAATTCATCAGGATTGA 1638  
|||||  
Db 680 AGCCA-CCACAC-GTATACC-ATGGTCCATGTCGCTATGGAATAGCC-ATGATCTTTA 735  
|||||  
Cp 1637 AGCCAACCCACACAGTATACCTATGATGTCGCTGATGGAATAGCCCATGATCTTTA 1578  
|||||  
Db 736 TATCTTT-AAACTCGCGTGCAGAAATCCCA-TGAGGATGTCGAGGCGCGGGGA 789  
|||||  
Cp 1577 TATCTTTAAACTCGCGTGCAGAAATCCCACTGAGGATGTCGAGGCGCGGGGA 1522  
|||||

RESULT 8  
LOCUS AW173232 653 bp mRNA EST 16-NOV-1999  
DEFINITION x185a05.x1 Soares\_NFL\_T\_GBC\_S1 Homo sapiens cDNA clone  
IMAGE:2663984 3' similar to gb:M58342 IDURONATE 2-SULFATASE  
PRECURSOR (HUMAN); mRNA sequence.  
AW173232  
ACCESSION AW173232.1 GI:6439180  
VERSION EST.  
KEYWORDS human.  
SOURCE Homo sapiens  
ORGANISM Homo sapiens  
REFERENCE Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;  
Eutheria; Primates; Catarrhini; Hominidae; Homo.  
AUTHORS 1 (bases 1 to 653)  
TITLE NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.  
JOURNAL National Cancer Institute, Cancer Genome Anatomy Project (CGAP),  
Tumor Gene Index  
COMMENT Unpublished (1997)  
On Apr 30, 1999 this sequence version replaced gi:5866661.  
Contact: Robert Strausberg, Ph.D.  
Tel.: (301) 496-1550  
Email: Robert.Strausberg@nih.gov  
This clone is available royalty-free through LLNL; contact the  
IMAGE Consortium (info@image.llnl.gov) for further information.  
Seq primer: -40up from Gibco  
High quality sequence stop: 440.  
Location/Qualifiers  
1. .653  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/clone="IMAGE:2663984"  
/clone\_lib="Soares\_NFL\_T\_GBC\_S1"  
/lab\_host="DH10B"  
/note="Organ: pooled; Vector: pT73D-Pac (Pharmacia) with  
a modified polylinker; Site\_1: Not I; Site\_2: Eco RI;  
Equal amounts of plasmid DNA from three normalized  
libraries (fetal lung NBHL19W, testis NHT, and B-cell  
NCI-CGAP-GCBI) were mixed, and ss circles were made in  
vitro. Following HAP purification, this DNA was used as  
tracer in a subtractive hybridization reaction. The driver

was PCR-amplified cDNAs from pools of 5,000 clones made  
from the same 3 libraries. The pools consisted of  
I.M.A.G.E. clones 297480-302087, 682632-687239,  
726408-728711, and 729096-731399. Subtraction by Bento  
Soares and M. Fatima Bonaldo.

BASE COUNT 143 a 196 c 154 g 157 t 3 others

Query Match 27.8%; Score 638; DB 50; Length 653;  
Best Local Similarity 98.9%; Pred. No. 0.00e+00;

Matches 647; Conservative 0; Mismatches 6; Indels 1; Gaps 1;

Db 1 GCTGGGCTGCTGCTGAGCTCCGCTCGCTCGCCCTCGGATCGGAAACGAGGCGCAACTC 60  
|||||

Qy 160 GCTGGGCTGCTGCTGAGCTCCGCTCGCTCGCCCTCGGATCGGAAACGAGGCGCAACTC 219  
|||||

Db 61 GACCACAGATGCTCTGAACGTTCTTCTCATCATCGTGGATGACCTGCGCCCTCCCTGGG 120  
|||||

Qy 220 GACCACAGATGCTCTGAACGTTCTTCTCATCATCGTGGATGACCTGCGCCCTCCCTGGG 279  
|||||

Db 121 CTGTTATGGGGATAAGCTGCTGAGTCCGCCAAATATTTGACCAACTGCCACAGCCT 180  
|||||

Qy 280 CTGTTATGGGGATAAGCTGCTGAGTCCGCCAAATATTTGACCAACTGCCACAGCCT 339  
|||||

Db 181 CCTCTTCCAGAAATGCTTTGCGCAGCAAGCAGTGTGCGCCCGGAGCGCGTTCTTTCCT 240  
|||||

Qy 340 CCTCTTCCAGAAATGCTTTGCGCAGCAAGCAGTGTGCGCCCGGAGCGCGTTCTTTCCT 399  
|||||

Db 241 CACTGCGAGGAGACCTTGACACACCCGCCCTGTACGACTTCAACTCTCTTGGAGGTGCA 300  
|||||

Qy 400 CACTGCGAGGAGACCTTGACACACCCGCCCTGTACGACTTCAACTCTCTTGGAGGTGCA 459  
|||||

Db 301 CGCTGGAATCTCTCCACCATCCCGCAGTACTTCAAGGAGAAATGGCTATGTGACCATGTC 360  
|||||

Qy 460 CGCTGGAATCTCTCCACCATCCCGCAGTACTTCAAGGAGAAATGGCTATGTGACCATGTC 519  
|||||

Db 361 GGTGGGAAAAGTCTTTCACCCCTGGGATATCTTCTAACCATACCGATGATCTCCGTATAG 420  
|||||

Qy 520 GGTGGGAAAAGTCTTTCACCCCTGGGATATCTTCTAACCATACCGATGATCTCCGTATAG 579  
|||||

Db 421 CTGGTCTTTTCCACCTTATCATCTTCTCTGAGAGTATGAAACACTAAGACATGTCG 480  
|||||

Qy 580 CTGGTCTTTTCCACCTTATCATCTTCTCTGAGAGTATGAAACACTAAGACATGTCG 639  
|||||

Db 481 AGGGCCAGATGGAGAACTCCATGCCAACCTGCTNTGCCCTGTGGATGTGTTGATGTTCC 540  
|||||

Qy 640 AGGGCCAGATGGAGAACTCCATGCCAACCTGCTNTGCCCTGTGGATGTGTTGATGTTCC 699  
|||||

Db 541 CGAGGCGACCTTGCCTGACAAACAGAGCACTGAGCAAGCCATACANGTGTGGAAAAGAT 600  
|||||

Qy 700 CGAGGCGACCTTGCCTGACAAACAGAGCACTGAGCAAGCCATACANGTGTGGAAAAGAT 759  
|||||

Db 601 GATAAGCTCAGCAGTCTTCTCTCTGCGCGGTGG-TATCANTAGCCACAT 653  
|||||

Qy 760 GAAAAGCTCAGCAGTCTTCTCTCTGCGCGGTGG-TATCANTAGCCACAT 813  
|||||

RESULT 9  
LOCUS AA553786 704 bp mRNA EST 12-AUG-1997  
DEFINITION n135d03.s1 NCI-CGAP\_Lul Homo sapiens cDNA clone IMAGE:978921 3'  
similar to gb:M58342 IDURONATE 2-SULFATASE PRECURSOR (HUMAN); mRNA  
sequence.

ACCESSION AA553786

VERSION AA553786.1 GI:2324325

KEYWORDS EST.

SOURCE human.

ORGANISM Homo sapiens

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;

Eutheria; Primates; Catarrhini; Hominidae; Homo.

REFERENCE 1 (bases 1 to 704)

AUTHORS NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.

TITLE National Cancer Institute, Cancer Genome Anatomy Project (CGAP),  
Tumor Gene Index

JOURNAL COMMENT

Unpublished (1997)  
On Apr 14, 1993 this sequence version replaced gi:693458.  
Contact: Robert Strausberg, Ph.D.  
Tel: (301) 496-1550  
Email: Robert.Strausberg@nih.gov  
Tissue Procurement: L. Jeffrey Medeiros, M.D., Michael R. Emmert-Buck, M.D., Ph.D.  
cDNA Library Preparation: Stratagene, Inc., David B. Krizman, Ph.D.  
cDNA Library Arraying: Greg Lennon, Ph.D.  
DNA Sequencing by: Washington University Genome Sequencing Center  
Clone distribution: NCI-CGAP clone distribution information can be found through the I.M.A.G.E. Consortium/LLNL at:  
www-bio.llnl.gov/bbrp/image/image.html

Seq primer: -40ml3 fwd. ET from Amersham  
High quality sequence stop: 392.

FEATURES source

1..704  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/clone="IMAGE:978821"  
/clone\_lib="NCI-CGAP\_Lu1"  
/tissue\_type="lung tumor"  
/lab\_host="SOLR (kanamycin resistant)"  
/note="Organ: lung; Vector: Bluescript SK-; Site\_1: EcoRI; Site\_2: XhoI; Cloned unidirectionally. Primer: Oligo dT. Bulk lung tumor. 5' adaptor sequence: 5' GAATTCGGCAGCAG 3' 3' adaptor sequence: 5' CTCGAGTCTTTTCTTTTCTTTT 3'  
Average insert size: 1.1 kb."  
211 a 163 c 161 g 169 t

BASE COUNT ORIGIN

Query Match 27.8%; Score 639; DB 19; Length 704;  
Best Local Similarity 96.9%; Pred. No. 0.00e+00;  
Matches 686; Conservative 0; Mismatches 17; Indels 5; Gaps 5;

Db 1 ACCAAATTATAAATTTTAAATAAGACTAAACGAAAGGTTTGGCTGTACATATTCTCAGG 60  
|||||  
Cp 1980 ACCAATTATAAATTTTAAATAAGACTAAACGAAAGGTTTGGCTGTACATATTCTCAGG 1921  
|||||  
Db 61 CCAAAATTTGATGCATGTTTGGATTAACTAGCCCTCAGGCTGCTTCCAAATATTATGGGT 120  
|||||  
Cp 1920 CCAAAATTTGATGCATGTTTGGATTAACTAGCCCTCAGGCTGCTTCCAAATATTATGGGT 1861  
|||||  
Db 121 AATCACAAGGACCGCTCTAATCTCTCTCCTCAGGCTGGAGGAGGACATCAGAT 180  
|||||  
Cp 1860 AATCACAAGGACCGCTCTAATCTCTCTCCTCAGGCTGGAGGAGGACATCAGAT 1801  
|||||  
Db 181 TTGCCATCCATGGTTGGCAAAACTCAAGGCATCAACAACTGGAAGATCTCCACCTTGG 240  
|||||  
Cp 1800 TTGCCATCCATGGTTGGCAAAACTCAAGGCATCAACAACTGGAAGATCTCCACCTTGG 1741  
|||||  
Db 241 GAATCAATATACATATTGTGATCCAGCAATGGGTGAGATCCACAAATACAGATTCGCCCT 300  
|||||  
Cp 1740 GAATCAATATACATATTGTGATCCAGCAATGGGTGAGATCCACAAATACAGATTCGCCCT 1681  
|||||  
Db 301 GCATGGATGCAGAAAGTTAGCTAGAAAATTCATCAGATTGAAGCCCAACCCACAGTA 360  
|||||  
Cp 1680 GCATGGATGCAGAAAGTTAGCTAGAAAATTCATCAGATTGAAGCCCAACCCACAGTA 1621  
|||||  
Db 361 TACCTATAGTCTATGGTGGCGTATGGAATAGCCATGATCTTTATATCTTTTAAACTCGGC 420  
|||||  
Cp 1620 TACCTATAGTCTATGGTGGCGTATGGAATAGCCATGATCTTTATATCTTTTAAACTCGGC 1561  
|||||  
Db 421 TTGTTCAGAAATTCACCTGAGGATGCTCTGAAGCGCGGGATAGTACTGGCTATAGGCAATCAGT 480  
|||||  
Cp 1560 TTGTTCAGAAATTCACCTGAGGATGCTCTGAAGCGCGGGATAGTACTGGCTATAGGCAATCAGT 1501  
|||||  
Db 481 TCACGGGGAATTCACGGGAGGTACCGATCCTACTTCCAGTACAGGATCGAAATGCTT 540  
|||||  
Cp 1500 TCACGGGGA-TTACCAGGGAGGTACCGATCCTCTTCCAAAGTCACGGAATCGAAATGCTT 1442  
|||||

Db 541 CAGAGGTTCTTGGCTTCTGCAAGAGCTCAACGTGAATGAAGNACGGGCGAGG 600  
|||||  
Cp 1441 CAGAGGTTCTTGGCTTCTGCAAGAGCTCAACGTGAATGAAGNACGGGCGAGG 1382  
|||||  
Db 601 TGGACCC-GCAGTCTGCA-GTCCAGCCAGCTGGGAGAAAGACACAGTTTCCCA-G 657  
|||||  
Cp 1381 TGGAACTTGCAGTCTGCAAGTCCAGCCAGCTGGGAGAAAGACACAGTTTCCCAAG 1322  
|||||  
Db 658 GTCCAGGAGTTCCGCTGCTGCTCCATCAATGGTGAAGCG-AATCAAA 704  
|||||  
Cp 1321 GTCCATGATGCTGCTGCTGCTCCATCACTGAGGCGGAATCAAA 1274  
|||||

RESULT 10

LOCUS AA513794 605 bp mRNA EST 20-AUG-1997  
DEFINITION nh89c05.r1 NCI-CGAP\_Brl.1 Homo sapiens cDNA clone IMAGE:965672 5' similar to gb:M58342 IDURONATE 2-SULFATASE PRECURSOR (HUMAN);, mRNA sequence.

ACCESSION

AA513794

VERSION

AA513794.1 GI:2252215

KEYWORDS

EST.

SOURCE

human.

ORGANISM

Homo sapiens

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

REFERENCE

1 (bases 1 to 606)

AUTHORS

NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.

TITLE

National Cancer Institute, Cancer Genome Anatomy Project (CGAP), Tumor Gene Index  
Unpublished (1997)  
On Nov 29, 1993 this sequence version replaced gi:636840.  
Contact: Robert Strausberg, Ph.D.  
Tel: (301) 496-1550  
Email: Robert.Strausberg@nih.gov  
Tissue Procurement: Christopher Moskaluk, M.D., Ph.D., Michael R. Emmert-Buck, M.D., Ph.D.  
cDNA Library Preparation: M. Bento Soares, Ph.D.  
cDNA Library Arrayed by: Greg Lennon, Ph.D.  
DNA Sequencing by: Washington University Genome Sequencing Center  
Clone distribution: NCI-CGAP clone distribution information can be found through the I.M.A.G.E. Consortium/LLNL at:  
www-bio.llnl.gov/bbrp/image/image.html

FEATURES

source

Insert Length: 2236 Std Error: 0.00  
Seq primer: -28ml3 rev1 ET from Amersham  
High quality sequence stop: 476.  
Location/Qualifiers  
1..606  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/clone="IMAGE:965672"  
/clone\_lib="NCI-CGAP\_Brl.1"  
/sex="female, pooled"  
/tissue\_type="breast"  
/lab\_host="DH10B"  
/note="vector: pT7T3D-Pac (Pharmacia) with a modified polylinker; 1st strand cDNA was prepared from pooled bulk breast tumor tissue, and was then primed with a Not I - oligo(dT) primer. Double-stranded cDNA was ligated to Eco RI adaptors (Pharmacia), digested with Not I and cloned into the Not I and Eco RI sites of the modified pT7T3 vector. Library is not normalized. (The normalized version of this library is NCI-CGAP\_Br2.) Library was constructed by Bento Soares and M. Fatima Bonaldo."

BASE COUNT 126 a 188 c 146 g 144 t 2 others

ORIGIN

Query Match 25.7%; Score 590; DB 18; Length 606;

Best Local Similarity 98.8%; Pred. No. 0.00e+00;

Matches 600; Conservative 0; Mismatches 6; Indels 1; Gaps 1;

Db 1 CCCGACCGCGGAGGC-TTCTCTGGCTGGTCTGGTCTGAGTCCGTCGTCGCC 59  
|||||

QY 135 CCCGACGGCCGAGGCGCTTCTCTGGCTGGGCTGGTCTGAGCTCCGTCCTCGTGGCCCC 194

Db 60 TCGGATCGGAACGAGCGCCAACTCGACACAGATGCTCTGAACGTCTTCTTCATCATCG 119

QY 195 TCGGATCGGAACGAGCGCCAACTCGACACAGATGCTCTGAACGTCTTCTTCATCATCG 254

Db 120 TGGATGACCTCGGCCCCCTCCCTGGGCTGTATGGGATAAGCTGTGAGGTCGCCAATA 179

QY 255 TGGATGACCTCGGCCCCCTCCCTGGGCTGTATGGGATAAGCTGTGAGGTCGCCAATA 314

Db 180 TTGACCAACTGGCATCCACAGCCTCTTTCCCAAGTCCCTTTCGCGACACAGGAGTGT 239

QY 315 TTGACCAACTGGCATCCACAGCCTCTCTTCCAGAAATGCCCTTTCGCGACAGGAGTGT 374

Db 240 GCGCCCCGAGCGGCTTTCTTCTCACTGCGCAGGAGACCTGACACACCCGCTGTAGC 299

QY 375 GCGCCCCGAGCGGCTTTCTTCTCACTGCGCAGGAGACCTGACACACCCGCTGTAGC 434

Db 300 ACTTCAACTCTACTGAGGGTGCACGCTGGAACCTTCTCCACCATCCGCCAGTACTTCA 359

QY 435 ACTTCAACTCTACTGAGGGTGCACGCTGGAACCTTCTCCACCATCCGCCAGTACTTCA 494

Db 360 AGGAGATGCTATGTGACCATGTGGGTGGGAAAGTCTTTACCCCTGGGATATCTTCA 419

QY 495 AGGAGATGCTATGTGACCATGTGGGTGGGAAAGTCTTTACCCCTGGGATATCTTCA 554

Db 420 ACCATACCGATGATCTCCGTATAGCTGGTCTTTTCCACCTTATCATCTTCTCTGAGA 479

QY 555 ACCATACCGATGATCTCCGTATAGCTGGTCTTTTCCACCTTATCATCTTCTCTGAGA 614

Db 480 AGTATGAACACATGAAGACATGTCCAGGCGCAGATGGAGACTCCATGCCAACCTGCTTT 539

QY 615 AGTATGAACACATGAAGACATGTCCAGGCGCAGATGGAGACTCCATGCCAACCTGCTTT 674

Db 540 GCCCTGTGGATGTGCTGGATGTTCCGAGGCGACCTTCCCTGACAAACAGACACTGAGC 599

QY 675 GCCCTGTGGATGTGCTGGATGTTCCGAGGCGACCTTCCCTGACAAACAGACACTGAGC 734

Db 600 AAGCCAT 606

QY 735 AAGCCAT 741

RESULT 11

LOCUS AAL131455 644 bp mRNA EST 14-MAY-1997

DEFINITION z132d04.r1 Soares\_pregnant\_uterus\_NbHPU Homo sapiens cDNA clone IMAGE:503623 5', similar to gb:M58342 IDURONATE 2-SULFATASE

PRECUSOR (HUMAN);, mRNA sequence.

ACCESSION AAL131455

VERSION AAL131455.1 GI:1692996

KEYWORDS EST.

SOURCE human.

ORGANISM Homo sapiens

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.

REFERENCE 1 (bases 1 to 644)

AUTHORS Hallier,L., Lennon,G., Becker,M., Bonaldo,M.F., Chiapelli,B., Chissone,S., Dietrich,N., DuBuque,T., Favello,A., Gish,W., Hawkins,M., Hultman,M., Kucaba,T., Lacy,M., Le,M., Le,N., Mardis,E., Moore,B., Morris,M., Parsons,J., Prange,C., Rifkin,L., Rohlfing,T., Schellenberg,K., Soares,M.B., Tan,F., Thierry-Mieg,J., Trevasakis,E., Underwood,K., Wohlmann,P., Waterston,R., Wilson,R. and Marra,M.

TITLE Generation and analysis of 280,000 human expressed sequence tags

JOURNAL Genome Res. 6 (9), 807-828 (1996)

MEDLINE 97044478

COMMENT On Nov 29, 1993 this sequence version replaced gi:503248. Contact: Wilson RK Washington University School of Medicine 4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108 Tel: 314 286 1800 Fax: 314 286 1810 Email: est@watson.wustl.edu

This clone is available royalty-free through LNL ; contact the IMAGE Consortium (info@image.llnl.gov) for further information. Insert Length: 849 Std Error: 0.00 Seq primer: -28M13 rev2 from Amersham High quality sequence stop: 490.

FEATURES

Location/Qualifiers

source

1..644

/organism="Homo sapiens"

/db\_xref="GDB:3808611"

/db\_xref="taxon:9606"

/clone="IMAGE:503623"

/clone\_lib="Soares\_pregnant\_uterus\_NbHPU"

/sex="female"

/dev\_stage="adult"

/lab\_host="DH10B"

/note="Organ: uterus; Vector: pT73-Pac; Site\_1: Not I; Site\_2: Eco RI; 1st strand cDNA was primed with a Not I - oligo(dT) primer [5', AACTGGAGAATTCGGCGCGCTTTTTTTTTTTTTTTT 3'], double-stranded cDNA was ligated to Eco RI adaptors (Pharmacia), digested with Not I and cloned into the Not I and Eco RI sites of the modified pT73 vector. Library went through one round of normalization. Library constructed by M. Fatima Bonaldo."

BASE COUNT 147 a 156 c 152 g 183 t 6 others

ORIGIN

Query Match 25.0%; Score 574; DB 13; Length 644;

Best Local Similarity 96.9%; Pred. No. 0.00e+00;

Matches 624; Conservative 0; Mismatches 13; Indels 7; Gaps 7;

Db 1 GGTGGGCTCTAGTGAACATGGAGATGGGCGCAATATACAGCAATTTTGATGTGCTACCC 60

QY 1131 GGTGGGCTCTAGTGAACATGGAGATGGGCGCAATATACAGCAATTTTGATGTGCTACCC 1190

Db 61 ATCTTCCCTCGATATCTATGTTCTTCCGAAAGGAGCGGCTTCACATCCCGAGGAGCGGAGA 120

QY 1191 ATGTTCCCTCGATATCTATGTTCTTCCGAAAGGAGCGGCTTCACATCCCGAGGAGCGGAGA 1250

Db 121 AGCTTTTCCCTTACCTCGACCCCTTTTGATTTCCGCGCTCACAGTTGATGAGCGAGGCGAGC 180

QY 1251 AGCTTTTCCCTTACCTCGACCCCTTTTGATTTCCGCGCTCACAGTTGATGAGCGAGGCGAGC 1310

Db 181 AATCCATGGACCTTGTGGAACTTGTCTCTTTTCCACGCTGGCTGGACTTCAGGAC 240

QY 1311 AATCCATGGACCTTGTGGAACTTGTCTCTTTTCCACGCTGGCTGGACTTCAGGAC 1370

Db 241 TGCAGGTTCCACATCGCTGCCGCTTCTTCATTTCCAGTTGAGCTGTGAGAGAGGCA 300

QY 1371 TGCAGGTTCCACATCGCTGCCGCTTCTTCATTTCCAGTTGAGCTGTGAGAGAGGCA 1430

Db 301 AGAACCTTCTGAAGACATTTTTCGATTCGCTGACTTGGAGAGGATCCGCTACCTCCCTGGTA 360

QY 1431 AGAACCTTCTGAAGACATTTTCGATTCGCTGACTTGGAGAGGATCCGCTACCTCCCTGGTA 1490

Db 361 ATCCCGTGAACATGATTCCTATAGCCAGTATCCCGGCGCTTCAGACATCCCTCAGTGA 420

QY 1491 ATCCCGTGAACATGATTCCTATAGCCAGTATCCCGGCGCTTCAGACATCCCTCAGTGA 1550

Db 421 ATTCTGACAGCGAGTGTAAAAGATATAAGATCATGGGCTATTCCTCAGGACCATAG 480

QY 1551 ATTCTGACAGCGAGTGTAAAAGATATAAGATCATGGGCTATTCCTCAGGACCATAG 1610

Db 481 ACTATAGTNTACTGTCTGGTGGCTTCAATCCCTCATGAATNTCAGTCACTTTCT 540

QY 1611 ACTATAGTNTACTGTCTGGTGGCTTCAATCCCTCATGAATNTCAGTCACTTTCT 1669

Db 541 GACCATTCAGCAGGGGAAACGTATTTTGGTGGAAATTCCTGACCATTCAGGATNCACAA 600

QY 1670 GAC-ATCCATGCAGGGGAACTGTATTTTG-TGGA-TTCTGACCATTCAGGAT-CACAA 1725

Db 601 TATGTAGNATGAATTTCCCAAGGTTGAAGATCTTTTCNAGTTGT 644

QY 1726 TATGTATAATGA-TTCCCAAGGTGGA-CATCTTTTCCAGTTGTT 1767

RESULT 12  
LOCUS AI478887 629 bp mRNA EST 14-APR-1999  
DEFINITION tm25b02.x1 Soares\_NFL\_T\_GBC\_S1 Homo sapiens cDNA clone  
IMAGE:2157579.3' similar to gb:M58342 IDURONATE 2-SULFATASE  
PRECURSOR (HUMAN);, mRNA sequence.  
ACCESSION AI478887  
VERSION AI478887.1 GI:4372055  
KEYWORDS EST.  
SOURCE human.  
ORGANISM Homo sapiens  
REFERENCE Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;  
AUTHORS Eutheria; Primates; Catarrhini; Hominidae; Homo.  
TITLE NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.  
Tumor Gene Index  
JOURNAL National Cancer Institute, Cancer Genome Anatomy Project (CGAP),  
COMMENT Unpublished (1997)  
On Jun 5, 1998 this sequence version replaced gi:3187940.  
Contact: Robert Strausberg, Ph.D.  
Tel: (301) 496-1550  
Email: Robert.Strausberg@nih.gov  
This clone is available royalty-free through LLNL ; contact the  
IMAGE Consortium (info@image.llnl.gov) for further information.  
Insert Length: 812 Std Error: 0.00  
Seq primer: -400P from Gibco  
High quality sequence stop: 453.  
Location/Qualifiers  
1. .629  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/clone="IMAGE:2157579"  
/clone.lib="Soares\_NFL\_T\_GBC\_S1"  
/lab\_host="DH10B"  
/note="Organ: pooled; Vector: pTT3D-Pac (Pharmacia) with  
a modified polylinker; Site\_1: Not I; Site\_2: Eco RI;  
Equal amounts of plasmid DNA from three normalized  
libraries (fetal lung NBHL19W, testis NHT, and B-cell  
NCI-CGAP\_GCB1) were mixed, and ss circles were made in  
vitro. Following HAP purification, this DNA was used as  
tracer in a subtractive hybridization reaction. The driver  
was PCR-amplified cDNAs from pools of 5,000 clones made  
from the same 3 libraries. The pools consisted of  
I.M.A.G.E. clones 297480-302087, 682632-687239,  
726408-728711, and 729096-731399. Subtraction by Bento  
Soares and M. Fatima Bonaldo. "  
BASE COUNT 215 a 101 c 110 g 202 t 1 others  
ORIGIN

Query Match 24.5%; Score 563; DB 32; Length 629;  
Best Local Similarity 97.8%; Pred. No. 0.00e+00;  
Matches 590; Conservative 0; Mismatches 10; Indels 3; Gaps 3;

Db 13 GAAAAGAAACATGGATTACATGTTAACTAATAATTGTAATATATAGGGCTCATAAGAT 72  
|||||  
Cp 2296 GAAAAGAAACATGGATTACATGTTAACTAATAATTGTAATATATAGGGCTCATAAGAT 2237  
|||||  
Db 73 ATATTATTAATTAATAACAAAATAAATCAGGAACCTTAGTGTGGTGCCTAGTTTGATATAT 132  
Cp 2236 ATATTATTAATTAATAACAAAATAAATCAGGAACCTTAGTGTGGTGCCTAGTTTGATATAT 2177  
|||||  
Db 133 GATTACTTTTGAATGCCTAAATCCACAATAAATGAAGATATCTTTGTGTATATG 192  
|||||  
Cp 2176 GATTACTTTTGAATGCCTAAATCCACAATAAATGAAGATATCTTTGTGTATATG 2117  
|||||  
Db 193 TTATGTTTGGTTATGATGCTTCGTATCCAAAGGTATGACATAACTTTGAGTTGT 252  
|||||  
Cp 2116 TTATGTTTGGTTATGATGCTTCGTATCCAAAGGTATGACATAACTTTGAGTTGT 2057  
|||||  
Db 253 TTGCTTTGTTATTTTATTCAGTAAATAAGCCGTAACCTGTTTAAAAAGAGGAAATTAATAA 312  
|||||

Cp 2056 TTGCTTTGTTATTTATTCAGTAATAAGCCGTAACTGTTTTTAAAAAGAGGAAATTAATAA 1997  
Db 313 AAAA-CTGGTCCAAATACCAATATATAAATTTTAATAAGACTAAACAGAGGTTTGGCT 371  
|||||  
Cp 1996 AAAAATCTGGTCCAAATACCAATATATAAATTTTAATAAGACTAAACAGGAAAGTTTGGCT 1937  
Db 372 GTTACATATTCTCAGGCCAAATTTGTTGATGCATGTTTGGATTAACTAGCCCTCAGGCTGC 431  
|||||  
Cp 1936 GTTACATATTCTCAGGCCAAATTTGTTGATGCATGTTTGGATTAACTAGCCCTCAGGCTGC 1877  
Db 432 TTCCCAATATTATGGGTAATCACAACAGCAGCTCTAACTCTCTCTCTCACCAGCTGGA 491  
|||||  
Cp 1876 TTCCCAATATTATGGGTAATCACAACAGCAGCTCTAACTCTCTCTCACCAGCTGGA 1917  
Db 492 AGGAGCACATCACATTTTCCCATCCATGTTGGCAAACTCAAGGCATCAACACACTGGGA 551  
|||||  
Cp 1816 AGGAGCACATCACATTTTCCCATCCATGTTGGCAAACTCAAGGCATCAACACACTGGGA 1757  
Db 552 AAGATCTTCACCTTGGGATCATTTATACATTTGTGATC-TGCAATGG-TCAGAAGTCAC 609  
|||||  
Cp 1756 AAGATCTTCACCTTGGGATCATTTATACATTTGTGATCTGCAATGGTCAAGAATCCAC 1697  
Db 610 AAA 612  
Cp 1696 AAA 1694

RESULT 13  
LOCUS AA236900 576 bp mRNA EST 03-MAR-1997  
DEFINITION zs43h10.s1 Soares\_NHMPu\_S1 Homo sapiens cDNA clone IMAGE:688003 3'  
similar to gb:M58342 IDURONATE 2-SULFATASE PRECURSOR (HUMAN);, mRNA  
sequence.  
ACCESSION AA236900  
VERSION AA236900.1 GI:1860930  
KEYWORDS EST.  
SOURCE human.  
ORGANISM Homo sapiens  
REFERENCE Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;  
AUTHORS Eutheria; Primates; Catarrhini; Hominidae; Homo.  
1 (bases 1 to 576)  
Hillier,L., Clark,N., Dubuque,T., Elliston,K., Hawkins,M.,  
Holman,M., Hultman,M., Kucaba,T., Le.M., Lennon,G., Marra,M.,  
Parsons,J., Rifkin,L., Rohlfing,T., Soares,M., Tan,F.,  
Trevasaki,E., Waterston,R., Williamson,A., Wohldmann,P. and  
Wilson,R.  
TITLE The WashU-Merck EST Project  
JOURNAL Unpublished (1995)  
COMMENT On Sep 12, 1996 this sequence version replaced gi:1397523.  
Contact: Wilson RK  
Washington University School of Medicine  
4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108  
Tel: 314 286 1800  
Fax: 314 286 1810  
Email: est@watson.wustl.edu  
This clone is available royalty-free through LLNL ; contact the  
IMAGE Consortium (info@image.llnl.gov) for further information.  
Seq primer: -41m13 fwd. Et from Amersham  
High quality sequence stop: 306.  
Location/Qualifiers  
1. .576  
/organism="Homo sapiens"  
/db\_xref="GDB:5592159"  
/db\_xref="taxon:9606"  
/clone="IMAGE:688003"  
/clone.lib="Soares\_NHMPu\_S1"  
/tissue\_type="Pooled human melanocyte, fetal heart, and  
pregnant uterus"  
/lab\_host="DH10B"  
/note="Organ: mixed (see below); Vector: pTT3D-Pac  
(Pharmacia) with a modified polylinker; Site 1: Not I;  
Site 2: Eco RI; Equal amounts of plasmid DNA from three  
normalized libraries (melanocyte 2NBHM, pregnant uterus  
NBPU, and fetal heart NBHL19W) were mixed, and ss circles

were made in vitro. Following HAP purification, this DNA was used as tracer in a subtractive hybridization reaction. The driver was PCR-amplified cDNAs from pools of 5,000 clones made from the same 3 libraries. The pools consisted of I.M.A.G.E. clones 260232-265223, 340488-343479, and 484488-489479."

BASE COUNT	175 a	130 c	119 g	152 t
ORIGIN				
Query Match	24.4%	Score 560;	DB 14;	Length 576;
Best Local Similarity	99.0%;	Pred. No. 0.00e+00;		
Matches	571;	Conservative	0;	Mismatches 5; Indels 1; Gaps 1;
Db	1	ACCAATTATAATTTTATAAAGACTAAACGAAAGGTTTGCTGTGTACATATTCACAGG	60	
Cp	1980	ACCAATTATAATTTTATAAAGACTAAACGAAAGGTTTGCTGTGTACATATTCACAGG	1921	
Db	61	CCAAATTTGTCATGTCATGTTGGATTAACTAGCCCTCAGGCTGCTTCCCAATATATATGGGT	120	
Cp	1920	CCAAATTTGTCATGTCATGTTGGATTAACTAGCCCTCAGGCTGCTTCCCAATATATATGGGT	1861	
Db	121	AATCACAAAACGACGCTCTAACTCTCTCTCACCAGCTGGAAGGAGACATACAT	180	
Cp	1860	AATCACAAAACGACGCTCTAACTCTCTCTCACCAGCTGGAAGGAGACATACAT	1801	
Db	181	TGCGCATCCATGGTTGGCAAACTCAAGGCATCAACAACCTGGAAGAGATCTCCACCTGG	240	
Cp	1800	TGCGCATCCATGGTTGGCAAACTCAAGGCATCAACAACCTGGAAGAGATCTCCACCTGG	1741	
Db	241	GAATCATTTATACATATTTGATCTCTGCAATGGGTGAGATCCAGATCCACAAAATACAGTTC	300	
Cp	1740	GAATCATTTATACATATTTGATCTCTGCAATGGGTGAGATCCAGATCCACAAAATACAGTTC	1681	
Db	301	GCATGGATGTCAGAAAAGTTAGCTAGAAATTCATCAGGATTTGAAGCCAAACCCACACAGTA	360	
Cp	1680	GCATGGATGTCAGAAAAGTTAGCTAGAAATTCATCAGGATTTGAAGCCAAACCCACACAGTA	1621	
Db	361	TACCTATAGTCTATGGTCGGTATGAATAGCCCATGATCTTTATATCTTTTAACTCGGC	420	
Cp	1620	TACCTATAGTCTATGGTCGGTATGAATAGCCCATGATCTTTATATCTTTTAACTCGGC	1561	
Db	421	TTGTGAGAAATCCACTGAGGGATGCTGAAGCCGGGGATAGTGGCTATAGGCAATCACT	480	
Cp	1560	TTGTGAGAAATCCACTGAGGGATGCTGAAGCCGGGGATAGTGGCTATAGGCAATCACT	1501	
Db	481	TCACGGGATTTACAGGGAGGTACGGATTTCTTCCGAGTACGGAATCGAAAATGCTTC	540	
Cp	1500	TCACGGGATTTACAGGGAGGTACGGATTTCTTCCGAGTACGGAATCGAAAATGCTTC	1441	
Db	541	AGAAGGTTCTTGACCTCTCTGCACAG-TCAACCTGAA	576	
Cp	1440	AGAAGGTTCTTGACCTCTCTGCACAGTCAACGTGAA	1404	

RESULT	14	A1769955	566 bp	mrna	EST	28-JUN-1999
LOCUS						
DEFINITION	WJ22f11.x1 NCI-CCAP_Kid12 Homo sapiens cDNA clone IMAGE:2403597 3' similar to gb:M58342 IDURONATE 2-SULFATASE PRECURSOR (HUMAN);, mRNA sequence.					
ACCESSION	A1769955.1					
VERSION	EST.					
KEYWORDS	EST.					
SOURCE	human.					
ORGANISM	Homo sapiens					
REFERENCE	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.					
AUTHORS	1 (bases 1 to 566)					
TITLE	NCI-CCAP http://www.ncbi.nlm.nih.gov/ncicgap. National Cancer Institute, Cancer Genome Anatomy Project (CGAP), Tumor Gene Index					
JOURNAL	Unpublished (1997)					
COMMENT	On May 18, 1998 this sequence version replaced gi:3136710. Contact: Robert Strausberg, Ph.D.					

Tel: (301) 496-1550  
Email: Robert.Strausberg@nih.gov  
Tissue Procurement: Christopher Moskaluk, M.D., Ph.D., Michael R. Emmert-Buck, M.D., Ph.D.  
CDNA Library Preparation: M. Bento Soares, Ph.D.  
DNA sequencing by: Greg Lennon, Ph.D.  
Clone distribution: NCI-CCAP clone distribution information can be found through the I.M.A.G.E. Consortium/LLNL at: www.bio.llnl.gov/bbrp/image/image.html

Seq primer: -40UP from Gibco  
High quality sequence stop: 385.

FEATURES  
source

1..566  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/clone="IMAGE:2403597"  
/clone\_lib="NCI-CCAP\_Kid12"  
/tissue\_type="2 pooled tumors (clear cell type)"  
/lab\_host="DH10B"

/note="Organ: kidney; Vector: pT73D-Pac (Pharmacia) with a modified polylinker; Site\_1: Not 1; Site\_2: Eco RI; Plasmid DNA from the normalized library NCI-CCAP\_Kid5 was prepared, and ss circles were made in vitro. Following HAP purification, this DNA was used as tracer in a subtractive hybridization reaction. The driver was PCR-amplified cDNAs from a pool of 5,000 clones made from the same library (cloneIDs 132912-1325831, 1471368-1472903 and 1492104-1493255). Subtraction by Bento Soares and M. Fatima Bonaldo."

BASE COUNT 119 a 181 c 131 g 135 t  
ORIGIN

Query Match 24.0%; Score 552; DB 38; Length 566;  
Best Local Similarity 98.8%; Pred. No. 0.00e+00;  
Matches 559; Conservative 0; Mismatches 7; Indels 0; Gaps 0;

Db	1	TCGAAGCGGAAATGCCGCCACCCCGGACCGGTGCGAGCCCTTCTGCTGGCTGGTTC	60	
QY	114	TCGAAGCGGAAATGCCGCCACCCCGGACCGGTGCGAGCCCTTCTGCTGGCTGGTTC	173	
Db	61	TGAGCTCCGCTGCTGGTCCGCTCGGATCCGAACGAGGCGCACTCCACACAGATGCTC	120	
QY	174	TGAGCTCCGCTGCTGGTCCGCTCGGATCCGAACGAGGCGCACTCCACACAGATGCTC	233	
Db	121	TGAACGTTCTTCATCATCGTGCATGACCTCGGCCCTCCCTGGGCTGTATGGGATA	180	
QY	234	TGAACGTTCTTCATCATCGTGCATGACCTCGGCCCTCCCTGGGCTGTATGGGATA	293	
Db	181	AGCTGGTGGGTCGCCCAATATTGACCAACTGGCATCCACAGCTCCTCTCCAGAGATG	240	
QY	294	AGCTGGTGGGTCGCCCAATATTGACCAACTGGCATCCACAGCTCCTCTCCAGAGATG	353	
Db	241	CTTTGCGCAGCAGTGTGCGCCCGGACCGCGCTTTCTTTCTCACTGGCAGGAGAC	300	
QY	354	CTTTGCGCAGCAGTGTGCGCCCGGACCGCGCTTTCTTTCTCACTGGCAGGAGAC	413	
Db	301	CTGACACCCCGCTGTACGACTTCAACTCCTACTGGAGGGTCCAGCTGGAACCTCT	360	
QY	414	CTGACACCCCGCTGTACGACTTCAACTCCTACTGGAGGGTCCAGCTGGAACCTCT	473	
Db	361	CCACCATCCCGCTGTACTTCAAGGAGAAATGGCTATGTGACCATCTCGTGGGAAAGTCT	420	
QY	474	CCACCATCCCGCTGTACTTCAAGGAGAAATGGCTATGTGACCATCTCGTGGGAAAGTCT	533	
Db	421	TTCAACCTGGGATATCTTTCAACCATACCGATGATTTCTCGTATAGTGTCTATTCCAC	480	
QY	534	TTCAACCTGGGATATCTTTCAACCATACCGATGATTTCTCGTATAGTGTCTATTCCAC	593	
Db	481	CTTATCATCTCTCTGTGAGAAAGTATGAAACACCTAGACATGTCGATGGCCAAATGG	540	
QY	594	CTTATCATCTCTCTGTGAGAAAGTATGAAACACCTAGACATGTCGATGGCCAAATGG	653	



Db 541 AACTCCATGCCAACCTGCTTTGCCCT 566  
|||||  
QY 654 AACTCCATGCCAACCTGCTTTGCCCT 679

RESULT 15 AI418603 551 bp mRNA EST 09-MAR-1999  
LOCUS t937d08.x1 Soares\_NFL\_T\_GBC\_s1 Homo sapiens cDNA clone  
DEFINITION IMAGE:2110959 3' similar to gb:M58342 IDURONATE 2-SULFATASE  
PRECURSOR (HUMAN); mRNA sequence.  
ACCESSION AI418603  
VERSION AI418603.1 GI:4264534  
KEYWORDS EST.  
SOURCE human.  
ORGANISM Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;  
Eutheria; Primates; Catarrhini; Hominidae; Homo.  
1 (bases 1 to 551)  
REFERENCE NCI-CGAP <http://www.ncbi.nlm.nih.gov/ncicgap>.  
AUTHORS National Cancer Institute, Cancer Genome Anatomy Project (CGAP),  
TITLE Tumor Gene Index  
JOURNAL Unpublished (1997)  
COMMENT On Mar 10, 1998 this sequence version replaced gi:2948359.  
Contact: Robert Strausberg, Ph.D.  
Tel: (301) 496-1550  
Email: Robert.Strausberg@nih.gov  
This clone is available royalty-free through LLNL; contact the  
IMAGE Consortium ([info@image.llnl.gov](mailto:info@image.llnl.gov)) for further information.  
Seq primer: -400P from Gibco  
High quality sequence stop: 453.  
Location/Qualifiers  
1..551  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/clone="IMAGE:2110959"  
/lab\_host="DH10B"  
/notes="Organ: pooled; Vector: pT7T3D-Pac (Pharmacia) with  
a modified polylinker; Site 1: Not 1; Site 2: Eco RI;  
Equal amounts of plasmid DNA from three normalized  
libraries (fetal lung NBHL19W, testis NHT, and B-cell  
NCI-CGAP\_GCB1) were mixed, and ss circles were made in  
vitro. Following HAP purification, this DNA was used as  
tracer in a subtractive hybridization reaction. The driver  
was PCR-amplified cDNAs from pools of 5,000 clones made  
from the same 3 libraries. The pools consisted of  
I.M.A.G.E. clones 297480-302087, 682632-687239,  
726408-728711, and 729096-731399. Subtraction by Bento  
Soares and M. Fatima Bonaldo."

FEATURES  
source

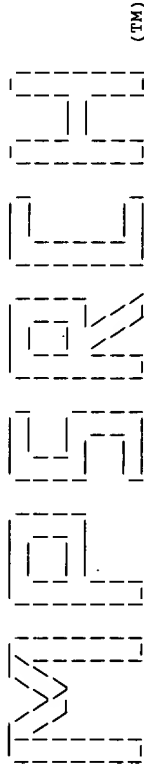
BASE COUNT 116 a 176 c 128 g 131 t  
ORIGIN  
Query Match 23.9%; Score 549; DB 31; Length 551;  
Best Local Similarity 99.8%; Pred. No. 0.00e+00;  
Matches 550; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Db 1 AATGCGGCACCGCGGACCGCGGAGCCCTTCTCTGGTGGGTCTGGTTCTGAGCTCCG 60  
|||||  
QY 123 AAATGCGGCACCGCGGACCGCGGAGCCCTTCTCTGGTGGGTCTGGTTCTGAGCTCCG 182  
|||||  
Db 61 TCTGCGTCGCGCTCGGATCCGAACGAGCGGCCACTCGACACAGATGCTCTGAACGTTT 120  
|||||  
QY 183 TCTGCGTCGCGCTCGGATCCGAACGAGCGGCCACTCGACACAGATGCTCTGAACGTTT 242  
|||||  
Db 121 TTCTCATCATCGTGGATGACCTGGCGCCCTCCCTGGGCTGTATGGGATAAGCTGTGA 180  
|||||  
QY 243 TTCTCATCATCGTGGATGACCTGGCGCCCTCCCTGGGCTGTATGGGATAAGCTGTGA 302  
|||||  
Db 181 GGTCGCCAAATATGACCAATGSCATCCACAGCCCTCCTCTCCAGAATGCCCTTTGCGC 240  
|||||  
QY 303 GGTCCGCCAAATATGACCAATGSCATCCACAGCCCTCCTCTCCAGAATGCCCTTTGCGC 362  
|||||

Db 241 AGCAAGCAGTGTGCGCGCGGAGCGGTTTCTTCTCACTGGCAGGAGACCTGACACCA 300  
|||||  
QY 363 AGCAAGCAGTGTGCGCGCGGAGCGGTTTCTTCTCACTGGCAGGAGACCTGACACCA 422  
|||||  
Db 301 CCGGCTGTACGACTTCAACTCCTACTGGAGGTGCACGCTGGAACCTTCTCCACCATCC 360  
|||||  
QY 423 CCGGCTGTACGACTTCAACTCCTACTGGAGGTGCACGCTGGAACCTTCTCCACCATCC 482  
|||||  
Db 361 CCCAGTACTTCAAGGAGAAATGGCTATGTGACCATGTCTGGTGGGAAAAGTCTTTTCACCCCTG 420  
|||||  
QY 483 CCCAGTACTTCAAGGAGAAATGGCTATGTGACCATGTCTGGTGGGAAAAGTCTTTTCACCCCTG 542  
|||||  
Db 421 TGATATCTTCTAACCATACCGGATGATCTCCGTTATAGCTGGTCTTTTCCACCTTATCATC 480  
|||||  
QY 543 GGATATCTTCTAACCATACCGGATGATCTCCGTTATAGCTGGTCTTTTCCACCTTATCATC 602  
|||||  
Db 481 CTTCTCTCTGAGAAATGATGAAACACTTAAGACATGTGAGGCGCAGATGAGAACTCCATG 540  
|||||  
QY 603 CTTCTCTCTGAGAAATGATGAAACACTTAAGACATGTGAGGCGCAGATGAGAACTCCATG 662  
|||||  
Db 541 CCAACCTGCTT 551  
|||||  
QY 663 CCAACCTGCTT 673

Search completed: Thu Apr 20 18:21:41 2000  
Job time : 7638 secs.

\*\*\*\*\*



\*\*\*\*\*

Release 3.1A John F. Collins, Biocomputing Research Unit.  
Copyright (c) 1993-1998 University of Edinburgh, U.K.  
Distribution rights by Oxford Molecular Ltd

MPSrch\_nn n.a. - n.a. database search, using Smith-Waterman algorithm  
Run on: Thu Apr 20 18:22:00 2000; MasPar time 483.16 Seconds  
Tabular output not generated. 1150.687 Million cell updates/sec

Title: >US-09-249-003-1  
Description: (1-2297) from US09249003.seq  
Perfect Score: 2297  
N.A. Sequence: 1 CGGCTGTGTCGCCAGTCTT.....TAATCATGTTCTTTTTC 2297  
Comp: GCCGACACACGCCGTCAGAA.....ATTAGGTACAAAGAAAGG

Scoring table: TABLE default  
Gap 6  
Mismatch STD : Dbase 0; Query 0  
Searched: 311545 seqs, 121019393 bases x 2  
Post-processing: Minimum Match 0%  
Listing first 45 summaries  
Database: n-geneseq36  
1:geneseqn  
Statistics: Mean 9.877; Variance 5.713; scale 1.729

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	ID	Description	Pred. No.
1	2297	100.0	2297	Human IDS CDNA.	0.00e+00
2	2297	100.0	2297	Human iduronate-2-sulf	0.00e+00
3	996	43.4	4428	Human IDS gene.	0.00e+00
4	996	43.4	4428	Human iduronate-2-sulf	0.00e+00
5	231	10.1	233	Human genome fragment	6.01e-140
6	48	2.1	204	Base substituted E.col	1.92e-13
7	48	2.1	204	Base substituted E.col	1.92e-13
8	45	2.0	91	Oligonucleotide probe	1.24e-11
9	40	1.7	91	Oligonucleotide probe	1.09e-08
10	37	1.6	501	Sequence encoding new	5.70e-07
11	37	1.6	501	Human adenosine Al rec	5.70e-07
12	35	1.5	114	Generic DNA sequence t	7.55e-06
13	34	1.5	114	Generic DNA sequence t	2.70e-05
14	35	1.5	178	Human endothelin-1 ant	7.55e-06
15	35	1.5	190	Chymase antisense olig	7.55e-06
16	35	1.5	190	Chymase antisense olig	7.55e-06
17	35	1.5	498	Sequence encoding new	7.55e-06
18	35	1.5	501	Sequence encoding new	7.55e-06
19	35	1.5	501	Sequence encoding new	7.55e-06
20	35	1.5	501	Sequence encoding new	7.55e-06

21	35	1.5	501	1	N50032	Sequence encoding new	7.55e-06
22	35	1.5	501	1	N50028	Sequence encoding new	7.55e-06
23	35	1.5	501	1	N50031	Sequence encoding new	7.55e-06
24	35	1.5	530	1	X54739	Human endothelin-1 ant	7.55e-06
25	35	1.5	567	1	X54764	Chymase antisense olig	7.55e-06
26	35	1.5	6225	1	X55273	Human enzyme-related a	7.55e-06
27	33	1.4	91	1	V44650	Mammalian DNA replicat	9.50e-05
28	33	1.4	91	1	V44650	Mammalian DNA replicat	9.50e-05
29	33	1.4	114	1	Q70469	Generic DNA sequence t	9.50e-05
30	32	1.4	114	1	Q70465	Generic DNA sequence t	3.30e-04
31	32	1.4	114	1	Q70470	Generic DNA sequence t	3.30e-04
32	32	1.4	114	1	Q70467	Generic DNA sequence t	3.30e-04
33	32	1.4	114	1	Q70466	Generic DNA sequence t	3.30e-04
34	33	1.4	501	1	N50030	Sequence encoding new	9.50e-05
35	33	1.4	501	1	N50029	Sequence encoding new	9.50e-05
36	33	1.4	7803	1	X55272	Human receptor-related	9.50e-05
37	31	1.3	114	1	Q70465	Generic DNA sequence t	1.13e-03
38	31	1.3	172	1	T76363	Human interleukin 8 an	1.13e-03
39	31	1.3	196	1	X54717	Human IL-8 antisense o	1.13e-03
40	31	1.3	266	1	X54711	Human IL-8 antisense o	1.13e-03
41	31	1.3	501	1	N50024	Sequence encoding new	1.13e-03
42	31	1.3	600	1	V86185	EST clone J635.	1.13e-03
43	31	1.3	600	1	V86185	EST clone J635.	1.13e-03
44	31	1.3	7036	1	X55274	Human factor-related a	1.13e-03
45	31	1.3	114955	1	X53491	Human adenosine Al rec	1.13e-03

ALIGNMENTS

RESULT 1  
ID V15698 standard; CDNA; 2297 BP.  
AC V15698;  
DT 01-JUN-1998 (first entry)  
DE Human IDS CDNA.  
KW Human; iduronate 2-sulphatase; IDS; treatment;  
KW Hunter syndrome; ds.  
OS Homo sapiens.  
FH Key Location/Qualifiers  
FT CDS 125..1777  
FT FT /\*tag= a  
FT FT /product= IDS  
PN US5728381-A.  
PD 17-MAR-1998.  
PF 07-JUN-1995; 484493.  
PR 17-DEC-1992; US-991973.  
PR 13-NOV-1991; US-790362.  
PR 28-NOV-1994; US-345212.  
PR 07-JUN-1995; US-484493.  
PA (ANSO/) ANSON D S.  
PA (BIEL/) BIELICKI J.  
PA (CLEM/) CLEMENTS P R.  
PA (HOPW/) HOPWOOD J J.  
PA (MORR/) MORRIS C P.  
PA (OCCH/) OCCHIODORO T.  
PA (WILS/) WILSON P J.  
PI Anson DS, Bielicki J, Clements PR, Hopwood JJ, Morris CP,  
PI Occhiodoro T, Wilson PJ;  
DR WPI; 98-206530/18.  
DR P-PSDB; W47298.  
PT Treatment of iduronate 2-sulphatase deficiency - comprises  
PT administering recombinant iduronate 2-sulphatase  
PS Example 1; Columns 23-28; 53pp; English.  
CC The present sequence encodes human iduronate 2-sulphatase (IDS).  
CC IDS deficiency can be treated by administering a recombinant human  
CC IDS that is more highly glycosylated than the naturally occurring  
CC enzyme, useful in the treatment of Hunter syndrome. The recombinant  
CC IDS may be administered in 0.5 microg/kg to 20 mg/kg doses. The  
CC administration route is oral, intravenous, intraperitoneal,  
CC intramuscular, subcutaneous or intranasal. The recombinant IDS has  
CC better uptake properties and/or a longer half-life in vivo, and is  
CC thus more efficient than naturally glycosylated IDS.  
SQ Sequence 2297 BP; 553 A; 600 C; 510 G; 634 T;

Query Match	100.0%;	Score 2297;	DB 1;	Length 2297;
Best Local Similarity	100.0%;	Pred. No. 0.00e+00;		
Matches 2297;	Conservative 0;	Mismatches 0;	Indels 0;	Gaps 0;
1	CGGCTGTGTTGCGCAGTCTTTCATGGTTC	CGCAGAGGAGTCTCTGTGGCTGCGCGGC	60	
1	CGGCTGTGTTGCGCAGTCTTTCATGGTTC	CGCAGAGGAGTCTCTGTGGCTGCGCGGC	60	
61	TGCTAACTGCGCCACCTGCTCGACCTGT	CCCGCGCGCTCTGAAGCGCGCTCGAAGC	120	
61	TGCTAACTGCGCCACCTGCTCGACCTGT	CCCGCGCGCTCTGAAGCGCGCTCGAAGC	120	
121	CGAATGCGCGCACCCGCGGACCGCGGAG	CGCTTCTTGCTGGGTCTGGTTC	180	
121	CGAATGCGCGCACCCGCGGACCGCGGAG	CGCTTCTTGCTGGGTCTGGTTC	180	
181	CGTCTGGTGCCTCGGATCCGAACCGCAG	CGCAACTCGACACAGATGCTCTGAACGT	240	
181	CGTCTGGTGCCTCGGATCCGAACCGCAG	CGCAACTCGACACAGATGCTCTGAACGT	240	
241	TCCTTCATCATCTGATGACCTGCGCCCT	CCCTGGGCTGTATGGGATAAAGCTGGT	300	
241	TCCTTCATCATCTGATGACCTGCGCCCT	CCCTGGGCTGTATGGGATAAAGCTGGT	300	
301	GAGTCCCCCAAAATATGACCACTGGCAT	CCACAGCCCTCTTCCAGAAATGCCCTTGC	360	
301	GAGTCCCCCAAAATATGACCACTGGCAT	CCACAGCCCTCTTCCAGAAATGCCCTTGC	360	
361	GCACAAGCAGTGTGCGCCCGGAGCGCG	CTTCTTCTCACATGCGAGAGCTGACAC	420	
361	GCACAAGCAGTGTGCGCCCGGAGCGCG	CTTCTTCTCACATGCGAGAGCTGACAC	420	
421	CACCGCGCTGTAGACTTCAACTCCTACT	GGAGGTCGACGCTGGAACCTTCTCCACCA	480	
421	CACCGCGCTGTAGACTTCAACTCCTACT	GGAGGTCGACGCTGGAACCTTCTCCACCA	480	
481	CCCCAGTACTTCAAGGAGATGGCTATGT	GACCATGTGCGTGGGAAAAGTCTTTCA	540	
481	CCCCAGTACTTCAAGGAGATGGCTATGT	GACCATGTGCGTGGGAAAAGTCTTTCA	540	
541	TGGGATATCTTCAACCATACCGATGAT	TCTCCGTATAGTGTCTTTCCACCTTATCA	600	
541	TGGGATATCTTCAACCATACCGATGAT	TCTCCGTATAGTGTCTTTCCACCTTATCA	600	
601	TCCCTTCTCTGAGAAGTATGAAACACT	TAAGACATGTGCGAGGCCAGATGGAGAACT	660	
601	TCCCTTCTCTGAGAAGTATGAAACACT	TAAGACATGTGCGAGGCCAGATGGAGAACT	660	
661	TGCCAACCTGTTTGCCTGTGGATGTGCT	GGATGTTCCCGAGGGCACCTTGCCTGACAA	720	
661	TGCCAACCTGTTTGCCTGTGGATGTGCT	GGATGTTCCCGAGGGCACCTTGCCTGACAA	720	
721	ACAGAGCAGTGCAGCAAGCCATACAGT	GTGTTGGAAAGATGAAACGTCAGCCAGT	780	
721	ACAGAGCAGTGCAGCAAGCCATACAGT	GTGTTGGAAAGATGAAACGTCAGCCAGT	780	
781	CTTCCTGGCGGTTGGGTATCATAGCCAC	ATCCCTCTCAGATACCCCAAGGAATTTCA	840	
781	CTTCCTGGCGGTTGGGTATCATAGCCAC	ATCCCTCTCAGATACCCCAAGGAATTTCA	840	
841	GAAGTTGTATCCCTTGGGAACATCACCC	CTGGCCCGGATCCCGAGGTCCCTGATGGCT	900	
841	GAAGTTGTATCCCTTGGGAACATCACCC	CTGGCCCGGATCCCGAGGTCCCTGATGGCT	900	
901	ACCCCTGTGGCCTTACAAACCCCTGGAT	GGACATCAGCAGGAGACGTCCAAGCCCT	960	
901	ACCCCTGTGGCCTTACAAACCCCTGGAT	GGACATCAGCAGGAGACGTCCAAGCCCT	960	
961	AAACATCAGTGTGCCGTATGGTCCAAT	TCTGTGGACTTTCAGCGGAAATCCGCCAGAG	1020	
961	AAACATCAGTGTGCCGTATGGTCCAAT	TCTGTGGACTTTCAGCGGAAATCCGCCAGAG	1020	
1021	CTACTTTGCCCTGTGTGTCATATTTGG	ATACAGGTCGCGCCCTCTTGAGTGGCTTTGA	1080	

Qy	1021	CTACTTTGGCTCTGTGTCATATTTGGATACACAGTTCGCGCGCTCTTGAGTGGCTTTGGA	1080
Db	1081	CGATCTTACAGCTGGCCACACAGCACCATCATTTGATTAACCTGGGATCATGGGTGGCTCT	1140
Qy	1081	CGATCTTACAGCTGGCCACACAGCACCATCATTTGATTAACCTGGGATCATGGGTGGCTCT	1140
Db	1141	AGGTGAACATGAGAAATGGCCAAATACAGCAATTTGATTTGCTTACCATGTTCCCT	1200
Qy	1141	AGGTGAACATGAGAAATGGCCAAATACAGCAATTTGATTTGCTTACCATGTTCCCT	1200
Db	1201	GATATTCATGTTCTTGGAGGACGCGCTTCACTTCCGAGGAGCGGAGAGCTTTTCCC	1260
Qy	1201	GATATTCATGTTCTTGGAGGACGCGCTTCACTTCCGAGGAGCGGAGAGCTTTTCCC	1260
Db	1261	TTTACCTCGACCCCTTTTATTCCGCTCACAGTTGATGGAGCCAGGAGCAATCCATGGA	1320
Qy	1261	TTTACCTCGACCCCTTTTATTCCGCTCACAGTTGATGGAGCCAGGAGCAATCCATGGA	1320
Db	1321	CCCTGTGGAACCTTGTCTCTCTTTTCCACGCTGGCTGGACTTGCAGGACTGCAGGTTCC	1380
Qy	1321	CCCTGTGGAACCTTGTCTCTCTTTTCCACGCTGGCTGGACTTGCAGGACTGCAGGTTCC	1380
Db	1381	ACCTCGCTCCCGCTTCTTCAATTTTCCAGCTTGCTGAGCTGCAGAGGAGCAAGCACTTCT	1440
Qy	1381	ACCTCGCTCCCGCTTCTTCAATTTTCCAGCTTGCTGAGCTGCAGAGGAGCAAGCACTTCT	1440
Db	1441	GAAGCATTTTCCATTCCTGCTTGGAAAGGATCCGTACCTCCCTGGTAAATCCCGCTGA	1500
Qy	1441	GAAGCATTTTCCATTCCTGCTTGGAAAGGATCCGTACCTCCCTGGTAAATCCCGCTGA	1500
Db	1501	ACTGATTTGGCTATAGCCAGTATCCCGGCTTCAGACATCCCTCAGTGGAAATTTCTGACAA	1560
Qy	1501	ACTGATTTGGCTATAGCCAGTATCCCGGCTTCAGACATCCCTCAGTGGAAATTTCTGACAA	1560
Db	1561	GCCGAGTTTAAAGATATAAAGATCATGGCTTATCCATACGCAACATAGCATATAGGTA	1620
Qy	1561	GCCGAGTTTAAAGATATAAAGATCATGGCTTATCCATACGCAACATAGCATATAGGTA	1620
Db	1621	TACTGTGGGGTGGCTTCAATCCCTGATGAATTTCTAGCTAACTTTTCTGACATCCATGC	1680
Qy	1621	TACTGTGGGGTGGCTTCAATCCCTGATGAATTTCTAGCTAACTTTTCTGACATCCATGC	1680
Db	1681	AGGGAACTGTATTTTGTGGATTCGACCATTCAGGATCAACATATGTAATATGATTC	1740
Qy	1681	AGGGAACTGTATTTTGTGGATTCGACCATTCAGGATCAACATATGTAATATGATTC	1740
Db	1741	CCAAGTGGAGATCTTTTCCAGTGTGCTTGGCTTGGCTTTGCCAACCATGATGGCAA	1800
Qy	1741	CCAAGTGGAGATCTTTTCCAGTGTGCTTGGCTTGGCTTTGCCAACCATGATGGCAA	1800
Db	1801	ATGTGATGTGCTCCCTTCCAGCTGGTGAGAGGAGGATTTAGAGCTGGTCTGTTTGTGAT	1860
Qy	1801	ATGTGATGTGCTCCCTTCCAGCTGGTGAGAGGAGGATTTAGAGCTGGTCTGTTTGTGAT	1860
Db	1861	ACCCATAATTTTGGAGAGCAGCTGAGGCTAGTTAATCCAAACATGCATCAACAATTTGG	1920
Qy	1861	ACCCATAATTTTGGAGAGCAGCTGAGGCTAGTTAATCCAAACATGCATCAACAATTTGG	1920
Db	1921	CGTGAGAATATGTAAACAGCAAACTTTTTCGTTTGTAGTCTTTTAAATTTTAAATTTGT	1980
Qy	1921	CGTGAGAATATGTAAACAGCAAACTTTTTCGTTTGTAGTCTTTTAAATTTTAAATTTGT	1980
Db	1981	AATTTGGACAGTTTTTTTTTAAATTTCCCTCTTTTAAACAGTTTACGCTTATTTACTG	2040
Qy	1981	AATTTGGACAGTTTTTTTTTAAATTTCCCTCTTTTAAACAGTTTACGCTTATTTACTG	2040
Db	2041	AATAAATACAAAGCAAACTCAAGTTATGTCATACCTTTGGATACGAGGAGCAATACA	2100
Qy	2041	AATAAATACAAAGCAAACTCAAGTTATGTCATACCTTTGGATACGAGGAGCAATACA	2100
Db	2101	TAAATACCAAAACATTAACATTAACAAAGATCTTTTCATTTATTTGTGGAATTTAGTGC	2160
Qy	2101	TAAATACCAAAACATTAACATTAACAAAGATCTTTTCATTTATTTGTGGAATTTAGTGC	2160

```
Qy 2101 TAATAACCAACATACATATATACAAAGAAATACCTTTCATTATTTGTGGAATTTAGTGC 2160
Db 2161 ATTTCAAAAAGTAATCATATATCAAACTAGGACACACACATAAGTTCTCTGATATTTTGT 2220
Qy 2161 ATTTCAAAAAGTAATCATATATCAAACTAGGACACACACATAAGTTCTCTGATATTTTGT 2220
Db 2221 TATAATTTAATAATATATCTTATGAGCCCTATATATTTCAAAATATATGTTTAAACATGTA 2280
Qy 2221 TATAATTTAATAATATATCTTATGAGCCCTATATATTTCAAAATATATGTTTAAACATGTA 2280
Db 2281 TCCATGTTCTTTTCC 2297
Qy 2281 TCCATGTTCTTTTCC 2297

RESULT 2
ID V52836 standard; cDNA; 2297 BP.
AC V52836;
DE Human iduronate-2-sulfatase encoding cDNA.
KW Human; iduronate-2-sulfatase; IDS; liver; glycosylated; heparin sulphate;
KW dermatan sulphate; lysosome; Hunter syndrome; severe mental retardation;
OS Homo sapiens.
FH Location/Qualifiers
FT 125..177
FT /*tag= a
FT /product= *iduronate-2-sulfatase"
PN US5798239-A.
PD 25-AUG-1998.
PF 07-JUN-1995; 484494.
PR 17-DEC-1992; US-991973.
PR 12-NOV-1991; US-790362.
PR 28-NOV-1994; US-345212.
PR 07-JUN-1995; US-484494.
PA (WOMEN-) WOMEN'S & CHILDREN'S HOSPITAL.
PI Anson DS, Bielicki J, Clements PR, Hopwood JJ, Morris CP,
PI Occhiodoro T, Wilson PJ;
DR WPI; 98-480382/41.
DR P-PSDB; W69786.
PT Production of glycosylated iduronate-2-sulphatase enzyme (IDS) -
PT comprises culturing a host cell containing a nucleic acid encoding
PT the enzymatically active iduronate-2-sulphatase polypeptide
PS Example 1; Fig 1; 53pp; English.
CC A method has been developed for the production of glycosylated
CC iduronate-2-sulphatase enzyme (IDS). The method comprises culturing
CC a host cell containing a nucleic acid encoding the enzymatically
CC active IDS polypeptide where the host cell glycosylates the polypeptide
CC to a greater degree than a native IDS polypeptide expressed by a
CC natural human liver cell. The present sequence encodes human IDS
CC which is used in an example from the present invention. The
CC recombinant IDS is used to treat IDS deficiency where heparin sulphate
CC and dermatan sulphate accumulates in lysosomes resulting in Hunter
CC syndrome which is manifested by e.g. severe mental retardation,
CC skeletal deformities and stiff joints. The recombinant IDS possesses
CC inter alia an improved half life inter alia and improved uptake
CC properties in comparison to the naturally glycosylated molecule.
SQ Sequence 2297 BP; 553 A; 600 C; 510 G; 634 T;

Query Match 100.0%; Score 2297; DB 1; Length 2297;
Best Local Similarity 100.0%; Pred. No. 0.00e+00;
Matches 2297; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Db 1 CGGTGCTTGGCGAGTCTTCATGGTTCCTCCGACGAGGAGGTCTCTGTGGCTGCGGCGGC 60
Qy 1 CGGTGCTTGGCGAGTCTTCATGGTTCCTCCGACGAGGAGGTCTCTGTGGCTGCGGCGGC 60
Db 61 TGCTAACTGCGCCACCTGCTGAGCCTGTCCCGCGCGTCTCTGAAGCGCGCGGTGCGAAGC 120
Qy 61 TGCTAACTGCGCCACCTGCTGAGCCTGTCCCGCGCGTCTCTGAAGCGCGCGGTGCGAAGC 120
Db 121 CGAATGCGCGCCACCGCGGCGCGGAGGCTTCTCTGGCTGGGTCTGAGTCTGAGTCTC 180
Qy 121 CGAATGCGCGCCACCGCGGCGCGGAGGCTTCTCTGGCTGGGTCTGAGTCTGAGTCTC 180
```

```
Qy 121 CGAATGCGCGCCACCGCGGCGCGGAGGCTTCTCTGGCTGGGTCTGAGTCTGAGTCTC 180
Db 181 CGTCTGGCTCGCCCTCGGATCCGAAACGCGAGGCCAACTCGACACACAGATGCTCTGAACGT 240
Qy 181 CGTCTGGCTCGCCCTCGGATCCGAAACGCGAGGCCAACTCGACACACAGATGCTCTGAACGT 240
Db 241 TCTTCTCATCATCGTGGATGACCTGCGCCCTCCCTCGGCTGTTTATGGGATAAGCTGGT 300
Qy 241 TCTTCTCATCATCGTGGATGACCTGCGCCCTCCCTCGGCTGTTTATGGGATAAGCTGGT 300
Db 301 GAGTCCCCAAATATTTGACCAATGCGATCCCAAGCCTCTCTTCCAGAAATGCTTTTC 360
Qy 301 GAGTCCCCAAATATTTGACCAATGCGATCCCAAGCCTCTCTTCCAGAAATGCTTTTC 360
Db 361 GCAGCAAGCAGTGTGCGCCCGCGGCGGCTTCTTCTCACTGCGAGGAGACCTGACAC 420
Qy 361 GCAGCAAGCAGTGTGCGCCCGCGGCGGCTTCTTCTCACTGCGAGGAGACCTGACAC 420
Db 421 CACCGCGCTGTACGACTTCAACTCTTACTGGAGGCTGACGCTGGAACCTTCTCCACCAT 480
Qy 421 CACCGCGCTGTACGACTTCAACTCTTACTGGAGGCTGACGCTGGAACCTTCTCCACCAT 480
Db 481 CCCCCAGTACTTCAAGGAGAAATGGCTATGTGACCATGTGCGTGGGAAAGTCTTTTCAACC 540
Qy 481 CCCCCAGTACTTCAAGGAGAAATGGCTATGTGACCATGTGCGTGGGAAAGTCTTTTCAACC 540
Db 541 TGGGATATCTTTAAACCATACCGATGATTTCCGTATAGTGGTCTTTTCCACCTTATCA 600
Qy 541 TGGGATATCTTTAAACCATACCGATGATTTCCGTATAGTGGTCTTTTCCACCTTATCA 600
Db 601 TCCTTCTCTGAGAAAGTATGAAACACTAAGACATGTGCGAGGCGCAGATGGAGAACTCCA 660
Qy 601 TCCTTCTCTGAGAAAGTATGAAACACTAAGACATGTGCGAGGCGCAGATGGAGAACTCCA 660
Db 661 TGCCAACTGCTTTGCGCTGTGGATGTCTGGATGTTCGCCGAGGCGACCTTGCCTGACAA 720
Qy 661 TGCCAACTGCTTTGCGCTGTGGATGTCTGGATGTTCGCCGAGGCGACCTTGCCTGACAA 720
Db 721 ACAGAGCACTGAGCAAGCCATACAGTTGTTGGAAAGATGAAACGTCAGCAGTCTCTTT 780
Qy 721 ACAGAGCACTGAGCAAGCCATACAGTTGTTGGAAAGATGAAACGTCAGCAGTCTCTTT 780
Db 781 CTTCCTGGCCGTTGGGTATCATAGCCACACATCCCTTCAGATACCCCAAGAAATTTCA 840
Qy 781 CTTCCTGGCCGTTGGGTATCATAGCCACACATCCCTTCAGATACCCCAAGAAATTTCA 840
Db 841 GAAGTTGTATCCCTTGAGAAACATACCCCTGGCCCGCGATCCCGAGGTCTCTGATGCCCT 900
Qy 841 GAAGTTGTATCCCTTGAGAAACATACCCCTGGCCCGCGATCCCGAGGTCTCTGATGCCCT 900
Db 901 ACCCCCTGTGGCCTTACAACCCCTGGATGGACATCAGGCAACGCGGAAAGACGTCCTCAAGCCTT 960
Qy 901 ACCCCCTGTGGCCTTACAACCCCTGGATGGACATCAGGCAACGCGGAAAGACGTCCTCAAGCCTT 960
Db 961 AAACATCAGTGTGCGGTATGTCCTGAGTTCCTGAGATTTTCAGCGGAAATTCGCCAGAG 1020
Qy 961 AAACATCAGTGTGCGGTATGTCCTGAGTTCCTGAGTTCCTGAGTTCCTGAGTTCGCCAGAG 1020
Db 1021 CTACTTTGCCCTGTGTCATATTTGGATACACAGGTGCGCGCCTCTTGTAGTCTTTTGA 1080
Qy 1021 CTACTTTGCCCTGTGTCATATTTGGATACACAGGTGCGCGCCTCTTGTAGTCTTTTGA 1080
Db 1081 CGATCTTACGTCGGCAACAGCACCATCATTTGATTTTACCTCGGATCATGGTGGGCTCT 1140
Qy 1081 CGATCTTACGTCGGCAACAGCACCATCATTTGATTTTACCTCGGATCATGGTGGGCTCT 1140
Db 1141 AGGTGAACATGGAGAAATGGGCCAAATACAGCAATTTTGTATGTTGCTACCCATGTTCCCT 1200
Qy 1141 AGGTGAACATGGAGAAATGGGCCAAATACAGCAATTTTGTATGTTGCTACCCATGTTCCCT 1200
Db 1201 GATATCTATGTTCTTGGAGAGGCGCTTACCTTCCGGAGGCGGCGAGAGCTTTTCCC 1260
Qy 1201 GATATCTATGTTCTTGGAGAGGCGCTTACCTTCCGGAGGCGGCGAGAGCTTTTCCC 1260
```

Db 1261 TTACCTCGACCCCTTTTGATTCGCGCTCACAGTTGATGGAGCGCAGGCAATCCATGGA 1320  
QY 1261 TTACCTCGACCCCTTTTGATTCGCGCTCACAGTTGATGGAGCGCAGGCAATCCATGGA 1320  
Db 1321 CCTGTGGAACTGTGTCTCTTTTCCACCGCTGGCTGACCTTCAGGACTGCAGTTCC 1380  
QY 1321 CCTGTGGAACTGTGTCTCTTTTCCACCGCTGGCTGACCTTCAGGACTGCAGTTCC 1380  
Db 1381 ACCTCGTCGCCCTTCTTCATTTACGTTTGAGCTGTGCAGAGAAGCAAGAACCTTCT 1440  
QY 1381 ACCTCGTCGCCCTTCTTCATTTACGTTTGAGCTGTGCAGAGAAGCAAGAACCTTCT 1440  
Db 1441 GAAGCATTTTCGATTCGCTGACCTTGGAGAGGATCCGTACCTCCCTGGTAAATCCCGTGA 1500  
QY 1441 GAAGCATTTTCGATTCGCTGACCTTGGAGAGGATCCGTACCTCCCTGGTAAATCCCGTGA 1500  
Db 1501 ACTGATTCGCTATAGCCAGTATCCCGGCCCTTCAGACATCCCTCAGTGGAAATTCGACAA 1560  
QY 1501 ACTGATTCGCTATAGCCAGTATCCCGGCCCTTCAGACATCCCTCAGTGGAAATTCGACAA 1560  
Db 1561 GCCGAGTTTAAAGATATAAGATCATGGCTATTCCATAGCAGCACCATAGACTATAGTA 1620  
QY 1561 GCCGAGTTTAAAGATATAAGATCATGGCTATTCCATAGCAGCACCATAGACTATAGTA 1620  
Db 1621 TACTGTGTGGGTTGGCTTCAATCCTGATGAATTTCTAGCTAACTTTCTGACATCCCATGC 1680  
QY 1621 TACTGTGTGGGTTGGCTTCAATCCTGATGAATTTCTAGCTAACTTTCTGACATCCCATGC 1680  
Db 1681 AGGGAACTGTATTTTGTGATTCGACCCANTGCAGATCACAAATATGTAATGATTC 1740  
QY 1681 AGGGAACTGTATTTTGTGATTCGACCCANTGCAGATCACAAATATGTAATGATTC 1740  
Db 1741 CCAAGTGGAGACTTTTCCAGTTGTTGATGCCCTTGAGTTTGGCAACCATGGGACAA 1800  
QY 1741 CCAAGTGGAGACTTTTCCAGTTGTTGATGCCCTTGAGTTTGGCAACCATGGGACAA 1800  
Db 1801 ATGTGATGTGCTCCTTCCAGCTGGTGAGAGGAGGTAGAGCTGGTCTGTTTGTGAT 1860  
QY 1801 ATGTGATGTGCTCCTTCCAGCTGGTGAGAGGAGGTAGAGCTGGTCTGTTTGTGAT 1860  
Db 1861 ACCCATATATTTGAAGCAGCCTGAGGCTAGTTAATCCAAACATGCATCAACAAATTTGG 1920  
QY 1861 ACCCATATATTTGAAGCAGCCTGAGGCTAGTTAATCCAAACATGCATCAACAAATTTGG 1920  
Db 1921 CCTGAGAAATATGTAACAGCCAACTTTTCGTTTAGTCTTTTATTAATAATTTATATGTT 1980  
QY 1921 CCTGAGAAATATGTAACAGCCAACTTTTCGTTTAGTCTTTTATTAATAATTTATATGTT 1980  
Db 1981 AATTGGACCACTTTTAAATTTTAAATTTTAAATTTTAAACAGTTACGGCTTATTTACTG 2040  
QY 1981 AATTGGACCACTTTTAAATTTTAAATTTTAAATTTTAAACAGTTACGGCTTATTTACTG 2040  
Db 2041 AATAAATACAAAGCAACAACTCAAGTTATGTCATACCTTTGGATAGCAAGACCATACA 2100  
QY 2041 AATAAATACAAAGCAACAACTCAAGTTATGTCATACCTTTGGATAGCAAGACCATACA 2100  
Db 2101 TAATAACAAACATACATATATACAAAGAAATCTTTTCAATTTATTTGGAAATTTAGTGC 2160  
QY 2101 TAATAACAAACATACATATATACAAAGAAATCTTTTCAATTTATTTGGAAATTTAGTGC 2160  
Db 2161 ATTTCAAAAAGTATCATATATCAAACTAGGCACCACACTAAGTTCCTGATATTTTGT 2220  
QY 2161 ATTTCAAAAAGTATCATATATCAAACTAGGCACCACACTAAGTTCCTGATATTTTGT 2220  
Db 2221 TATAATTTAATAATATATCTTATGAGCCCTATATATCAAAATATATATGTTAAACATGAA 2280  
QY 2221 TATAATTTAATAATATATCTTATGAGCCCTATATATCAAAATATATATGTTAAACATGAA 2280  
Db 2281 TCCATGTTCTTTTCC 2297  
QY 2281 TCCATGTTCTTTTCC 2297

RESULT 3  
ID V15700 standard; DNA; 4428 BP.  
AC V15700;  
DT 01-JUN-1998 (first entry)  
DE Human IDS gene.  
KW Human; iduronate 2-sulphatase; IDS; treatment;  
KW Hunter syndrome; ds.  
OS Homo sapiens.  
PN US572831-A.  
PD 17-MAR-1998.  
PF 07-JUN-1995; 484493.  
PR 17-DEC-1992; US-991973.  
PR 12-NOV-1991; US-790362.  
PR 28-NOV-1994; US-345212.  
PR 07-JUN-1995; US-484493.  
PA (ANSO/) ANSON D S.  
PA (BIEL/) BIELICKI J.  
PA (CLEM/) CLEMENTS P R.  
PA (HOPW/) HOPWOOD J J.  
PA (MORR/) MORRIS C P.  
PA (OCCH/) OCCHIODORO T.  
PA (WILS/) WILSON P J.  
PI Anson DS, Bielicki J, Clements PR, Hopwood JJ, Morris CP,  
PI Occhiodoro T, Wilson PJ;  
DR WPI; 98-206530/18.  
PT Treatment of iduronate 2-sulphatase deficiency - comprises  
PT administering recombinant iduronate 2-sulphatase  
PS Example 1; Columns 33-38; 53pp; English.  
CC The present sequence is the human iduronate 2-sulphatase (IDS)  
CC gene. IDS deficiency can be treated by administering a recombinant  
CC human IDS that is more highly glycosylated than the naturally  
CC occurring enzyme, useful in the treatment of Hunter syndrome. The  
CC recombinant IDS may be administered in 0.5 microg/kg to 20 mg/kg  
CC doses. The administration route is oral, intravenous,  
CC intraperitoneal, intramuscular, subcutaneous or intranasal. The  
CC recombinant IDS has better uptake properties and/or a longer  
CC half-life in vivo, and is thus more efficient than naturally  
CC glycosylated IDS.  
SQ Sequence 4428 BP; 1033 A; 1064 C; 1042 G; 1285 T;  
Query Match 43.4%; Score 996; DB 1; Length 4428;  
Best Local Similarity 100.0%; Pred. No. 0.00e+00;  
Matches 996; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
Db 3433 CAGGAGGCAATCCATGGACCTTGGAACTTGTGTCTTTTCCACGCTGGCTGGAC 3492  
QY 1302 CAGGAGGCAATCCATGGACCTTGGAACTTGTGTCTTTTCCACGCTGGCTGGAC 1361  
Db 3493 TTGCAGGACTGCAGGTTCCACCTCGCTGCCCGTTCCCTTCATTTACGTTGAGCTGCA 3552  
QY 1362 TTGCAGGACTGCAGGTTCCACCTCGCTGCCCGTTCCCTTCATTTACGTTGAGCTGCA 1421  
Db 3553 GAGAAGGCAAGAACCTTCTGAAGCATTTTCGATTCGCTGACTTGGAGAGGATCCGTACC 3612  
QY 1422 GAGAAGGCAAGAACCTTCTGAAGCATTTTCGATTCGCTGACTTGGAGAGGATCCGTACC 1481  
Db 3613 TCCCTGGTAATCCCGTGAACCTTTCGATTCGCTATAGCCAGTATCCCGGCCCTTCAGACATCC 3672  
QY 1482 TCCCTGGTAATCCCGTGAACCTTTCGATTCGCTATAGCCAGTATCCCGGCCCTTCAGACATCC 1541  
Db 3673 CTCAGTGGAAATTCGACAGCCGAGTTTAAAGATATAAAGATCATGGGCTATTCATAC 3732  
QY 1542 CTCAGTGGAAATTCGACAGCCGAGTTTAAAGATATAAAGATCATGGGCTATTCATAC 1601  
Db 3733 GCACCATAGACTATAGGTATCTGTGGTTGGCTTCAATCTGATGAATTTCTAGCTA 3792  
QY 1602 GCACCATAGACTATAGGTATCTGTGGTTGGCTTCAATCTGATGAATTTCTAGCTA 1661  
Db 3793 ACTTTTCTGACATCCATGCAGGGGAACCTATTTTGTGGATTTCTGACCCATTCGAGGATC 3852  
QY 1662 ACTTTTCTGACATCCATGCAGGGGAACCTATTTTGTGGATTTCTGACCCATTCGAGGATC 1721

FT	exon	1481..1569	
FT		/*tag= h	
FT		/number= 4	
FT	intron	1570..1840	
FT		/*tag= i	
FT		/number= 4	
FT	exon	1841..2041	
FT		/*tag= j	
FT		/number= 5	
FT	intron	2042..2293	
FT		/*tag= k	
FT		/number= 5	
FT	exon	2294..2464	
FT		/*tag= l	
FT		/number= 6	
FT	intron	2465..2683	
FT		/*tag= m	
FT		/number= 6	
FT	exon	2684..2810	
FT		/*tag= n	
FT		/number= 7	
FT	intron	2811..3032	
FT		/*tag= o	
FT		/number= 7	
FT	exon	3033..3206	
FT		/*tag= p	
FT		/number= 8	
FT	intron	3207..3435	
FT		/*tag= q	
FT		/number= 8	
FT	exon	3436..3908	
FT		/*tag= r	
FT		/number= 9	
PN	US7982239-A.		
PD	25-AUG-1998.		
PF	07-JUN-1995;	484494.	
PR	17-DEC-1992;	US-991973.	
PR	12-NOV-1991;	US-790362.	
PR	28-NOV-1994;	US-345212.	
PR	07-JUN-1995;	US-484494.	
PA	(WOME-) WOMEN'S & CHILDREN'S HOSPITAL..		
PI	Anson DS, Bielicki J, Clements PR, Hopwood JU, Morris CP,		
PI	Ochiodoro T, Wilson PJ;		
DR	WPI: 98-480382/41.		
PT	Production of glycosylated iduronate-2-sulphatase enzyme (IDS) -		
PT	comprises culturing a host cell containing a nucleic acid encoding		
PT	the enzymatically active iduronate-2-sulphatase polypeptide		
PS	Example 1; Fig 7; 53pp: English.		
CC	A method has been developed for the production of glycosylated		
CC	iduronate-2-sulphatase enzyme (IDS). The method comprises culturing		
CC	a host cell containing a nucleic acid encoding the enzymatically		
CC	active IDS polypeptide where the host cell glycosylates the polypeptide		
CC	to a greater degree than a native IDS polypeptide expressed by a		
CC	natural human liver cell. The present sequence encodes human IDS		
CC	which is used in an example from the present invention. The		
CC	recombinant IDS is used to treat IDS deficiency where heparin sulphate		
CC	and dermatan sulphate accumulates in lysosomes resulting in Hunter		
CC	syndrome which is manifested by e.g. severe mental retardation,		
CC	skeletal deformities and stiff joints. The recombinant IDS possesses		
CC	inter alia an improved half life inter alia and improved uptake		
CC	properties in comparison to the naturally glycosylated molecule.		
SQ	Sequence 4428 BP; 1033 A; 1064 C; 1042 G; 1285 T;		
Query Match 43.4%; Score 996; DB 1; Length 4428;			
Best Local Similarity 100.0%; Pred. No. 0.00e+00;			
Matches 996; Conservative 0; Mismatches 0; Indels 0; Gaps 0;			
Db	3433	CAGCAGCAATCCATGGACCTGTGTGAACCTGTGTCTCTTTTCCACGCTGGCTGGAC	3492
Qy	1302	CAGCAGGCAATCCATGGACCTGTGTGAACCTGTGTCTCTTTTCCACGCTGGCTGGAC	1361
Db	3493	TTGCAGGACTGCAGGTTCACCTCGCTGCCCGTTCCTTCATTTCAGTTGAGCTGTGCA	3552





88

[illegible]





\*\*\*\*\*  
WIREH  
\*\*\*\*\*  
(TM)

Release 3.1A John F. Collins, Biocomputing Research Unit.  
Copyright (c) 1993-1998 University of Edinburgh, U.K.  
Distribution rights by Oxford Molecular Ltd

MPsrch\_nn n.a. - n.a. database search, using Smith-Waterman algorithm

Run on: Thu Apr 20 12:57:15 2000; MasPar time 5060.39 Seconds  
Tabular output not generated. 1476.553 Million cell updates/sec

Title: >US-09-249-003-1  
Description: (1-2297) from US09249003.seq  
Perfect Score: 2297  
N.A. Sequence: 1 CGGCTGTGTGGCGAGTCTT.....TAATCCAGTTTCTTTTCC 2297  
Comp: GCCGACACACGGCGTCAGAA.....ATTAGTACAAAGAAAAGG

Scoring table: TABLE default  
Gap 6

Nmatch STD : Dbase 0; Query 0

Searched: 197718 seqs, 1626453718 bases x 2

Post-processing: Minimum Match 0%  
Listing first 45 summaries

Database: emb161  
1:em\_ba1 2:em\_ba2 3:em\_fun 4:em\_htg1 5:em\_htg2 6:em\_htg3  
7:em\_hum1 8:em\_hum2 9:em\_hum3 10:em\_hum4 11:em\_in  
12:em\_om 13:em\_or 14:em\_ov 15:em\_pat 16:em\_pl 17:em\_ro  
18:em\_un 19:em\_vl  
genbank1-115  
20:gb\_ba1 21:gb\_ba2 22:gb\_htg1 23:gb\_htg2 24:gb\_htg3  
25:gb\_htg4 26:gb\_htg5 27:gb\_htg6 28:gb\_htg7 29:gb\_in1  
30:gb\_in2 31:gb\_om 32:gb\_ov

Statistics: Mean 13.051; Variance 5.508; scale 2.369

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES				
Result No.	Score	Query Match	Description	Pred. No.
C 1	373	16.2	13574 24 AC011389 Homo sapiens chromosom	0.00e+00
2	367	16.0	13574 24 AC011389 Homo sapiens chromosom	0.00e+00
3	75	3.3	16310 26 AC015390 Drosophila melanogaste	3.22e-37
4	75	3.3	113816 23 AC004358 Drosophila melanogaste	3.22e-37
5	40	1.7	1056 31 MVU87256 Mustela vison ct dinuc	2.73e-09
6	40	1.7	10772 30 AF012089 Drosophila melanogaste	2.73e-09
7	38	1.7	10772 30 AF012089 Drosophila melanogaste	2.73e-09
8	37	1.6	1056 31 MVU87256 Mustela vison ct dinuc	6.90e-08
9	34	1.5	56722 27 AC016207 Homo sapiens clone RP1	3.42e-05
C 10	34	1.5	56722 27 AC016207 Homo sapiens clone RP1	3.42e-05
C 11	34	1.5	59006 27 AC016220 Homo sapiens clone RP1	3.42e-05
C 12	35	1.5	67656 27 AC016439 Homo sapiens clone RP1	7.50e-06
C 13	34	1.5	176224 22 HSA175010 Homo sapiens chromosom	3.42e-05

14	34	1.5	190839 23	AL133350	Homo sapiens chromosom	3.42e-05
15	34	1.5	192798 26	AC015864	Homo sapiens chromosom	3.42e-05
C 16	32	1.4	121485 22	HSDJ812P3	Homo sapiens chromosom	6.60e-04
17	32	1.4	152878 29	CEV18D10A	Caenorhabditis elegans	6.60e-04
C 18	32	1.4	164587 24	AC009444	Homo sapiens clone 1_O	6.60e-04
C 19	33	1.4	192969 26	AC012359	Homo sapiens clone NHO	1.52e-04
C 20	32	1.4	198363 27	AC013545	Homo sapiens chromosom	6.60e-04
C 21	32	1.4	210296 23	HS1057B20	Homo sapiens chromosom	6.60e-04
22	32	1.4	247332 22	CEV18D10	Caenorhabditis elegans	6.60e-04
23	33	1.4	301690 28	AC010909	Homo sapiens clone RP1	1.52e-04
24	30	1.3	318 31	PR081266	Prionallurus bengalensis	1.14e-02
C 25	29	1.3	61613 24	AC008457	Homo sapiens chromosom	4.53e-02
C 26	30	1.3	97480 25	AC011993	Homo sapiens clone NHO	1.14e-02
27	31	1.3	102324 23	HS1168M15	Homo sapiens chromosom	2.79e-03
C 28	30	1.3	105525 22	AP000619	Homo sapiens chromosom	1.14e-02
29	30	1.3	126534 24	AC008750	Homo sapiens chromosom	1.14e-02
30	30	1.3	161428 24	AC009542	Homo sapiens chromosom	1.14e-02
31	30	1.3	165732 28	AC010794	Homo sapiens clone RP1	1.14e-02
32	31	1.3	166980 24	AC009756	Homo sapiens clone 45	2.79e-03
33	30	1.3	167886 28	AC009985	Homo sapiens clone RP1	1.14e-02
34	31	1.3	170953 24	AC008278	Homo sapiens clone NHO	2.79e-03
35	30	1.3	172200 24	AC007219	Homo sapiens chromosom	1.14e-02
C 36	31	1.3	174870 24	AC009786	Homo sapiens clone 44	2.79e-03
C 37	30	1.3	176781 26	AC015802	Homo sapiens chromosom	1.14e-02
C 38	30	1.3	180406 28	AC017004	Homo sapiens clone RP1	1.14e-02
39	31	1.3	182024 23	AL133258	Homo sapiens chromosom	2.79e-03
40	30	1.3	184062 22	AP000842	Homo sapiens chromosom	1.14e-02
C 41	30	1.3	187886 25	AC011851	Homo sapiens chromosom	1.14e-02
C 42	31	1.3	189045 28	AC010862	Homo sapiens clone RP1	2.79e-03
C 43	30	1.3	190768 22	HSJ190J20	Homo sapiens chromosom	1.14e-02
44	31	1.3	195874 27	AC011676	Homo sapiens clone RP1	2.79e-03
C 45	30	1.3	318663 22	HSDJ654A7	Homo sapiens chromosom	1.14e-02

ALIGNMENTS

1  
RESULT AC011389 13574 bp DNA HTG 06-OCT-1999  
LOCUS Homo sapiens chromosome 5 clone C1978SKB\_174D11, \*\*\* SEQUENCING IN  
DEFINITION PROGRESS \*\*\*, 1 unordered pieces.  
AC011389  
AC011389, 1 GI:6013552  
VERSION HTG; HTGS\_PHASE1.  
KEYWORDS human.  
SOURCE Homo sapiens  
ORGANISM Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;  
Eutheria; Primates; Catarrhini; Hominidae; Homo.  
REFERENCE 1 (bases 1 to 13574)  
AUTHORS DOE Joint Genome Institute.  
TITLE Sequencing of Human Chromosome 5  
JOURNAL Unpublished  
REFERENCE 2 (bases 1 to 13574)  
AUTHORS DOE Joint Genome Institute.  
JOURNAL Direct Submission  
TITLE Submitted (06-OCT-1999) Production Sequencing Facility, DOE Joint  
Genome Institute, 2800 Mitchell Drive, Walnut Creek, CA 94598, USA  
COMMENT www.jgi.doe.gov.  
\* NOTE: This is a 'working draft' sequence. It currently  
\* consists of 1 contigs. The true order of the pieces  
\* is not known and their order in this sequence record is  
\* arbitrary. Gaps between the contigs are represented as  
\* runs of N, but the exact sizes of the gaps are unknown.  
\* This record will be updated with the finished sequence  
\* as soon as it is available and the accession number will  
\* be preserved.  
\* 1 13574: contig of 13574 bp in length.  
\* Location/Qualifiers  
\* 1. 13574  
\* /organism="Homo sapiens"  
\* /gb\_xref="taxon:9606"  
\* /chromosome="5"  
\* /clone="C1978SKB\_174D11"  
BASE COUNT 3467 a 2760 c 115 g 462 t 6770 others







QY	1324	TGTGAACTTGTCCTCTTTTCCACGCTGGCTGGACTTGCAGGACTGCAGGTTCCACC	1383
Db	6946	DCAADADCBACADACACCHTRANGTHWTKRANGTHSDATANCBAADCBDBCADA	7005
QY	1384	TCGCTGCCCGCTCTCTTATTCATTCAGTGTGCAGAGAG-GCAAGAACCTTCTGA	1442
Db	7006	DCAAAADDDADDBAADBADABDCDDCCABDDCCCBBCBADCBDDAADCCADAA	7065
QY	1443	AGCATTTTCGATTC-GT-GACTTGGAG-AGGATCCCTAGCTCCTGGTATATCCCGTG	1499
Db	7066	CAACDAABAACDCAABDDDDDDDDDDDBDDDBDADCDCCDAABCAABDBDD	7125
QY	1500	-AAGTATGCTATAGCAGTATCCCGCTTCAGACATCCCTCAGTGAATCTGAC	1558
Db	7126	ADDAADAAACACACBBAADDDADAABDDCAABACCCBACBBBCACDDCAADA	7185
QY	1559	AAGCGAGTTTAAAGATATAAGATCATGGCTATTCCATA-CGCACCATAGACTATAG	1617
Db	7186	DDDBCCCCBBACDDCBACAAACCBDBCCACCCCBADADACBACBBACCCBBA	7245
QY	1618	GTATACGTGTGGGTGGCTTCAATCTCGATGAATTTCTAGTAACTTTCTGACATCCA	1677
Db	7246	BABCCBCDACAACDAACBCCBDBDBACCCBDBCBDBACBCCBDBBBAACBDCAC	7305
QY	1678	TGCAGGGGAACATGATTTTGTGGATCTGCACCATTCGAGGATCACAAT-ATGTATAATG	1736
Db	7306	DBACBBADBBACADDDADCDACBACBBAADDBADBDACBCCCBCCDCAACDDBCAD	7365
QY	1737	ATTCACCAAGTGGAGATCTTTTCCAGTTGTGTGATGCTTTGATGCTTGGCAACCATGGATG	1796
Db	7366	CBAADCBADDBCCCBACACADABDBBBBCCDCBDBADDBBDBCBADDBCB	7425
QY	1797	GCAATGTGTGTGCTCCCTCCAGCTGGTGAGAGGAGGTAGAGTGTGCTGTTTGT	1856
Db	7426	ADCBABBBACACABCBCCCBACBCCBDBCCBDBBDBBDBBDBBDBBDBBDBBDBB	7485
QY	1857	GATTACCATTAATTTGGAAGCAGCCTGAGGCTAGTTTAAATCCAAACATGCATCAACAAT	1916
Db	7486	ABDACABDA-ADBCBADDCCDDDBADCCDDDBBBDADADCDADDBDDCCDCDADCB	7544
QY	1917	TGCGCTCAGATATGTAACAGCCAAACCTTTCTGTTAGTCTTTTAAATTTATAT	1976
Db	7545	ABDACBDBACCBACBACBDBCCBBAACCCBDBCDADDDDBDDCDDBCDDBBA	7604
QY	1977	TGTAATTTGGACCAAGTTTTTTTTTAAATTTCCCTCTTTTAAACAGTTACGGCTATT	2036
Db	7605	ABADADDACACACBACBACBACBACBACBACBACBACBACBACBACBACBACBACB	7664
QY	2037	ACTGAATAAATACAA-AGCAAAACAACTCAAGTTATGTATACCTTTGGATACGA-AGA-	2093
Db	7665	BCDCBDAADAAACDCDCDD	7724
QY	2094	CCATACATAATACCAACATACATATACAAAGAACTTATATATTTTGGGAAT	2153
Db	7725	DBBBBBAADCCCBACACBDBBDBCCDCBDCBDCBDCBDCBDCBDCBDCBDCBDCB	7784
QY	2154	TTAGTGATTTCAAAAGTAATCATATATCAAACTAGGCACACACATAGTTCCGTGATTA	2213
Db	7785	BDABADCCACABDBABDBCCCBABDBABDBABDBABDBABDBABDBABDBABDBAB	7842
QY	2214	TTTTTTTATTAATTAATTAATATATATCTTTAGCGCCTATATATTTCAAAATATATGTTAA	2273
Db	7843	CADDBADDBADD	7855
QY	2274	CATGTAATCCATG	2286
RESULT	3		
LOCUS	AC015390	16310 bp	DNA
DEFINITION	Drosophila melanogaster, *** SEQUENCING IN PROGRESS ***	HTG	16-NOV-1999
ACCESSION	AC015390		pieces.
VERSION	AC015390.1	GI:6435945	

KEYWORDS	HTG; HTGS_PHASE2.
SOURCE	fruit fly.
ORGANISM	Drosophila melanogaster
REFERENCE	Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta; Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.
AUTHORS	1 (bases 1 to 16310)
TITLE	Adams, M. and Venter, J.C.
JOURNAL	Submitted (16-NOV-1999) Celera Genomics, 45 West Gude Drive, Rockville, MD, USA
COMMENT	This sequence was identified as CDM:10209831 by the submitter. For further information on this sequence e-mail to fly@celera.com. * NOTE: This is a 'working draft' sequence. * This sequence will be replaced * by the finished sequence as soon as it is available and * the accession number will be preserved.
FEATURES	Location/Qualifiers
source	1..16310
organism	/organism="Drosophila melanogaster"
db_xref	/db_xref="taxon:7227"
BASE COUNT	4280 a 3707 c 3728 g 4595 t
ORIGIN	
Query Match	3.3%; Score 75; DB 26; Length 16310;
Best Local Similarity	65.6%; Pred. No. 3,22e-37;
Matches	158; Conservative 0; Mismatches 83; Indels 0; Gaps 0;
Db	11701 ACATATCCCTGTGTGCACCCAGCGGAATTCGCTGCTCAGGGTTCGAAGACCGGATCTC 11760
QY	363 ACAGACATGTGCGCCCGGAGCGCGTTCTTCTCTCACTGGCAGGAGACCTGCACCA 422
Db	11761 TGCATCTGTACGACTTCTACAGTTACTGCGCACTTTCACCGGAATTTCCACCACTTTC 11820
QY	423 CCGCCTGTACGACTTCAACTCTCTGAGGGTGACCGTGGAACTTCTCCACCATCC 482
Db	11821 CGCAGTACTTCAAGGACGACGATATACACCTATAGCTGTGAAAGGTTTCCATCCTG 11880
QY	483 CCGAGTACTTCAAGGAGATGCTATGTACCATGTGCGTGGGAAAGTCTTTCACCTG 542
Db	11881 GCGTCTATCCAAACACGAGTACTATCCCTTTAGTTGGTCCGCTCCAGCAATTCGTC 11940
QY	543 GGATATCTTCAACCATACCGATGATTCCTCGGTATAGTGGTCTTTTCCACCTATCATC 602
Db	11941 C 11941
QY	603 C 603
RESULT	4
LOCUS	AC004358 113816 bp
DEFINITION	Drosophila melanogaster chromosome 3 clone DS01859 (D262) map 63C5-63D3 strain y; cn bw sp, *** SEQUENCING IN PROGRESS ***
ACCESSION	AC004358 AC003918 AC003919 AC003920 AC003921
VERSION	AC004358.3 GI:5656737
KEYWORDS	HTG; HTGS_PHASE1.
SOURCE	fruit fly.
ORGANISM	Drosophila melanogaster
REFERENCE	Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta; Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.
AUTHORS	1 (bases 1 to 113816)
COMMENT	Celniker, S.E., Aghayani, A., Arcaluna, T.T., Baxter, E., Blazej, R.G., Butenhoff, C., Champe, M., Chavez, C., Chew, M., Ciesiolka, L., Doyle, C.M., Farfan, D.E., Galle, R., George, R.A., Harris, N.L., Kim, E., Lee, B., Lewis, S., Li, P., Lomotan, M.A., Mazda, P., Moshrefi, A.R., Moshrefi, M., Nixon, K., Pacleb, J.M., Park, S., Pfeiffer, B., Poon, L., Sequeira, A., Sethi, H., Snir, E., Svirska, R., Wan, K.H., Weinburg, T., Zhang, R., Zieran, L.L. and Rubin, G.M.
TITLE	Sequencing of Drosophila melanogaster

JOURNAL  
REFERENCE  
AUTHORSUnpublished  
2 (bases 1 to 113816)

Celniker, S.E., George, R.A., Galle, R., Svirskas, R.R., Hoskins, R.A.,  
 Agbayani, A., Arcaina, T.T., Baxter, E., Blazej, R.G., Chavez, C.,  
 Chew, M., Doyle, C.M., Farfan, D.E., Flanagan, J., Houston, K.A.,  
 Hummasti, S.R., Karra, K., Kearney, L., Kim, S.H., Lee, B.,  
 Lomotan, M.A., Mak, J., Mazda, P., Mok, M.S., Moshrefi, A.R.,  
 Moshrefi, M., Nixon, K., Pacleb, J.M., Park, S., Pfeiffer, B., Punch, E.,  
 Snir, E., Twomey, B., Wan, K.H., Whitelaw, K.R., Yee, A., Zhang, R.,  
 Zieran, L.L. and Kimmel, B.E.

## Direct Submission

Submitted (07-MAR-1998) Drosophila Genome Center, Lawrence Berkeley  
 Laboratory, MS 64-121, Berkeley, CA 94720, USA

On Jul 30, 1999 this sequence version replaced gi:5597062.

For further information about this sequence, including its location  
 and relationship to other sequences, please visit our sequence  
 archive web site (<http://www.fruitfly.org/sequence/>) or send email  
 to [bdp@fruitfly.berkeley.edu](mailto:bdp@fruitfly.berkeley.edu). All contigs in this submission meet  
 the following cutoffs: length >= 200 bases. P1 library location:  
 20-35.

\* NOTE: This is a 'working draft' sequence. It currently  
 \* consists of 74 contigs. The true order of the pieces  
 \* is not known and their order in this sequence record is  
 \* arbitrary. Gaps between the contigs are represented as  
 \* runs of N, but the exact sizes of the gaps are unknown.  
 \* This record will be updated with the finished sequence.  
 \* as soon as it is available and the accession number will  
 \* be preserved.

1 524: contig of 524 bp in length  
 \* 525 604: gap of unknown length  
 \* 605 1219: contig of 615 bp in length  
 \* 1220 1299: gap of unknown length  
 \* 1300 1513: contig of 214 bp in length  
 \* 1514 1593: gap of unknown length  
 \* 1594 1943: contig of 350 bp in length  
 \* 1944 2023: gap of unknown length  
 \* 2024 2556: contig of 533 bp in length  
 \* 2557 2636: gap of unknown length  
 \* 2637 2927: contig of 291 bp in length  
 \* 2928 3007: gap of unknown length  
 \* 3008 3439: contig of 432 bp in length  
 \* 3440 3519: gap of unknown length  
 \* 3520 4133: contig of 614 bp in length  
 \* 4134 4213: gap of unknown length  
 \* 4214 4832: contig of 619 bp in length  
 \* 4833 4912: gap of unknown length  
 \* 4913 5973: contig of 1061 bp in length  
 \* 5974 6053: gap of unknown length  
 \* 6054 6511: contig of 458 bp in length  
 \* 6512 6591: gap of unknown length  
 \* 6592 13907: contig of 7316 bp in length  
 \* 13908 13987: gap of unknown length  
 \* 13988 22128: contig of 8141 bp in length  
 \* 22129 22208: gap of unknown length  
 \* 22209 46184: contig of 23976 bp in length  
 \* 46185 46264: gap of unknown length  
 \* 46265 86697: contig of 40433 bp in length  
 \* 86698 86777: gap of unknown length  
 \* 86778 87035: contig of 258 bp in length  
 \* 87036 87115: gap of unknown length  
 \* 87116 87321: contig of 206 bp in length  
 \* 87322 87401: gap of unknown length  
 \* 87402 87878: contig of 477 bp in length  
 \* 87879 87958: gap of unknown length  
 \* 87959 88359: contig of 401 bp in length  
 \* 88360 88439: gap of unknown length  
 \* 88440 89001: contig of 562 bp in length  
 \* 89002 89081: gap of unknown length  
 \* 89082 89486: contig of 405 bp in length  
 \* 89487 89566: gap of unknown length  
 \* 89567 89925: contig of 359 bp in length  
 \* 89926 90005: gap of unknown length  
 \* 90006 90390: contig of 385 bp in length

TITLE  
JOURNAL  
COMMENT

90391 90470: gap of unknown length  
 90471 90875: contig of 405 bp in length  
 90876 90955: gap of unknown length  
 90956 91405: contig of 450 bp in length  
 91406 91485: gap of unknown length  
 91486 91901: contig of 416 bp in length  
 91902 91981: gap of unknown length  
 91982 92394: contig of 413 bp in length  
 92395 92474: gap of unknown length  
 92475 92750: contig of 396 bp in length  
 92751 92950: gap of unknown length  
 92952 93412: contig of 462 bp in length  
 93413 93492: gap of unknown length  
 93493 93915: contig of 423 bp in length  
 93916 93995: gap of unknown length  
 93996 94222: contig of 227 bp in length  
 94223 94302: gap of unknown length  
 94303 94622: contig of 320 bp in length  
 94623 94702: gap of unknown length  
 94703 94941: contig of 239 bp in length  
 94942 95021: gap of unknown length  
 95022 95412: contig of 391 bp in length  
 95413 95492: gap of unknown length  
 95493 95816: contig of 324 bp in length  
 95817 95896: gap of unknown length  
 95897 96153: contig of 257 bp in length  
 96154 96233: gap of unknown length  
 96234 96683: contig of 450 bp in length  
 96684 96764: gap of unknown length  
 96765 97249: contig of 486 bp in length  
 97250 97329: gap of unknown length  
 97330 97676: contig of 347 bp in length  
 97677 97756: gap of unknown length  
 97757 98253: contig of 497 bp in length  
 98254 98333: gap of unknown length  
 98334 98884: contig of 551 bp in length  
 98885 98964: gap of unknown length  
 98965 99282: contig of 318 bp in length  
 99283 99362: gap of unknown length  
 99363 100019: contig of 657 bp in length  
 100020 100099: gap of unknown length  
 100100 100353: contig of 254 bp in length  
 100354 100433: gap of unknown length  
 100434 100993: contig of 560 bp in length  
 100994 101073: gap of unknown length  
 101074 101358: contig of 285 bp in length  
 101359 101438: gap of unknown length  
 101439 101851: contig of 413 bp in length  
 101852 101931: gap of unknown length  
 101932 102206: contig of 275 bp in length  
 102207 102286: gap of unknown length  
 102287 102567: contig of 281 bp in length  
 102568 102647: gap of unknown length  
 102648 103081: contig of 434 bp in length  
 103082 103161: gap of unknown length  
 103162 103605: contig of 444 bp in length  
 103606 103685: gap of unknown length  
 103686 104194: contig of 509 bp in length  
 104195 104274: gap of unknown length  
 104275 104672: contig of 398 bp in length  
 104673 104752: gap of unknown length  
 104753 105076: contig of 324 bp in length  
 105077 105156: gap of unknown length  
 105157 105494: contig of 338 bp in length  
 105495 105574: gap of unknown length  
 105575 105891: contig of 317 bp in length  
 105892 105971: gap of unknown length  
 105972 106391: contig of 420 bp in length  
 106392 106471: gap of unknown length  
 106472 106868: contig of 397 bp in length  
 106869 106948: gap of unknown length  
 106949 107593: contig of 645 bp in length  
 107594 107673: gap of unknown length









Db 35224 AATGATATTTTTT 35236  
 ||||| | |||||  
 QY 2283 CATGTCCTTTTTT 2295

```

RESULT 10
LOCUS AC016207 56722 bp DNA HTG 23-NOV-1999
DEFINITION Homo sapiens clone RP11-25H20, LOW-PASS SEQUENCE SAMPLING.
ACCESSION AC016207
VERSION AC016207.1 GI:6466587
KEYWORDS HTG: HTGS_PHASE0.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE
1 (bases 1 to 56722)
AUTHORS Birren,B., Linton,L., Nusbaum,C. and Lander,E.
TITLE Homo sapiens, clone RP11-25H20
JOURNAL Unpublished
REFERENCE
2 (bases 1 to 56722)
AUTHORS Birren,B., Linton,L., Nusbaum,C., Lander,E., Allen,N., Anderson,M.,
Baldwin,J., Barna,N., Beckerly,R., Boguslavskiy,L., Boukigalter,B.,
Brown,A., Castle,A., Colangelo,M., Collins,S., Collamore,A.,
Cooke,P., DeArelano,K., Dewar,K., Domino,M., Donelan,L., Doyle,M.,
Ferreira,P., FitzHugh,W., Forrest,C., Funke,R., Gage,D.,
Galagan,J., Gardyna,S., Grant,G., Hagos,B., Heaford,A., Horton,L.,
Howland,J.C., Johnson,R., Jones,C., Kann,L., Karatas,A., Klein,J.,
Lehoczky,J., Lieu,C., Locke,K., Macdonald,P., Marquis,N.,
McEwan,P., McGurk,A., McKernan,K., McLaughlin,J., Meldrim,J.,
Morrow,J., Naylor,J., Norman,C.H., O'Connor,T., O'Donnell,P.,
Peterson,K., Pollara,V., Riley,R., Roy,A., Santos,R., Severy,P.,
Stange-Thomann,N., Stojanovic,N., Subramanian,A., Talamas,J.,
Tsfaye,S., Tirrell,A., Vassiliev,H., Vo,A., Wheeler,J., Wu,X.,
Wyman,D., Ye,W.J., Zimmer,A. and Zody,M.
Direct Submission
Submitted (23-NOV-1999) Whitehead Institute/MIT Center for Genome
Research, 320 Charles Street, Cambridge, MA 02141, USA
All repeats were identified using RepeatMasker:
Snit, A.F.A. & Green, P. (1996-1997)
http://ftp.genome.washington.edu/RM/RepeatMasker.html
----- Genome Center
Center: Whitehead Institute/ MIT Center for Genome Research
Center code: WtBR
Web site: http://www-seq.wi.mit.edu
Contact: sequence_submission@genome.wi.mit.edu
----- Project Information
Center project name: L4574
Center clone name: 25_H_20
-----
* NOTE: This record contains 72 individual
* sequencing reads that have not been assembled into
* contigs. Runs of N are used to separate the reads
* and the order in which they appear is completely
* arbitrary. Low-pass sequence sampling is useful for
* identifying clones that may be gene-rich and allows
* overlap relationships among clones to be deduced.
* However, it should not be assumed that this clone
* will be sequenced to completion. In the event that
* the record is updated, the accession number will
* be preserved.
1 778: contig of 778 bp in length
gap of unknown length
779 1539: contig of 761 bp in length
gap of unknown length
1540 2321: contig of 782 bp in length
gap of unknown length
2322 3075: contig of 754 bp in length
gap of unknown length
3076 3885: contig of 810 bp in length
gap of unknown length
3886 4699: contig of 814 bp in length
gap of unknown length

```

```

* 4700 5482: contig of 783 bp in length
* gap of unknown length
* 5483 6252: contig of 770 bp in length
* gap of unknown length
* 6253 7020: contig of 768 bp in length
* gap of unknown length
* 7021 7844: contig of 824 bp in length
* gap of unknown length
* 7845 8678: contig of 834 bp in length
* gap of unknown length
* 8679 9498: contig of 820 bp in length
* gap of unknown length
* 9499 10298: contig of 800 bp in length
* gap of unknown length
* 10299 11082: contig of 784 bp in length
* gap of unknown length
* 11083 11865: contig of 783 bp in length
* gap of unknown length
* 11866 12638: contig of 773 bp in length
* gap of unknown length
* 12639 13435: contig of 797 bp in length
* gap of unknown length
* 13436 14219: contig of 784 bp in length
* gap of unknown length
* 14220 15013: contig of 794 bp in length
* gap of unknown length
* 15014 15763: contig of 750 bp in length
* gap of unknown length
* 15764 16528: contig of 765 bp in length
* gap of unknown length
* 16529 17296: contig of 768 bp in length
* gap of unknown length
* 17297 18069: contig of 773 bp in length
* gap of unknown length
* 18070 18855: contig of 786 bp in length
* gap of unknown length
* 18856 19641: contig of 786 bp in length
* gap of unknown length
* 19642 20497: contig of 856 bp in length
* gap of unknown length
* 20498 21280: contig of 783 bp in length
* gap of unknown length
* 21281 22064: contig of 784 bp in length
* gap of unknown length
* 22065 22848: contig of 784 bp in length
* gap of unknown length
* 22849 23639: contig of 791 bp in length
* gap of unknown length
* 23640 24419: contig of 780 bp in length
* gap of unknown length
* 24420 25193: contig of 774 bp in length
* gap of unknown length
* 25194 25972: contig of 779 bp in length
* gap of unknown length
* 25973 26744: contig of 772 bp in length
* gap of unknown length
* 26745 27520: contig of 776 bp in length
* gap of unknown length
* 27521 28261: contig of 741 bp in length
* gap of unknown length
* 28262 29052: contig of 791 bp in length
* gap of unknown length
* 29053 29842: contig of 790 bp in length
* gap of unknown length
* 29843 30623: contig of 781 bp in length
* gap of unknown length
* 30624 31411: contig of 788 bp in length
* gap of unknown length
* 31412 32225: contig of 814 bp in length
* gap of unknown length
* 32226 33057: contig of 832 bp in length
* gap of unknown length
* 33058 33839: contig of 782 bp in length

```







```

RESULT 12
LOCUS AC016439 67656 bp DNA HTG 26-NOV-1999
DEFINITION Homo sapiens clone RP11-30C21, LOW-PASS SEQUENCE SAMPLING.
ACCESSION AC016439
VERSION AC016439.1 GI:6468800
KEYWORDS HTG; HTGS_PHASE0.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 67656)
AUTHORS Birren,B., Linton,L., Nusbaum,C. and Lander,E.
TITLE Homo sapiens, clone RP11-30C21
JOURNAL Unpublished
AUTHORS
REFERENCE 2 (bases 1 to 67656)
AUTHORS Birren,B., Linton,L., Nusbaum,C., Lander,E., Allen,N., Anderson,M.,
Baldwin,J., Barna,N., Beckerly,R., Boguslavskiy,L., Boukhgalter,B.,
Brown,A., Castle,A., Colangelo,M., Collins,S., Collymore,A.,
Cooke,P., Dearellano,K., Dewar,K., Domino,M., Donelan,L., Doyle,M.,
Ferreira,P., FitzHugh,W., Forrest,C., Funke,R., Gage,D.,
Galagan,J., Gardyna,S., Grant,G., Hagos,B., Heaford,A., Horton,L.,
Howland,J.C., Johnson,R., Jones,C., Kann,L., Karatas,A., Klein,J.,
Lehoczky,J., Lieu,C., Locke,K., Macdonald,P., Marguis,N.,
McEwan,P., McGurk,A., McKernan,K., McLaughlin,J., Meldrim,J.,
Morrow,J., Naylor,J., Norman,C.H., O'Connor,T., O'Donnell,P.,
Peterson,K., Pollara,V., Riley,R., Roy,A., Santos,R., Severy,P.,
Stange-Thomann,N., Stojanovic,N., Subramanian,A., Talamas,J.,
Tesfaye,S., Tirrell,A., Vassiliev,H., Vo,A., Wheeler,J., Wu,X.,
Wyman,D., Ye,W.J., Zimmer,A. and Zody,M.
DIRECT SUBMISSION
Submitted (26-NOV-1999) Whitehead Institute/MIT Center for Genome
Research, 320 Charles Street, Cambridge, MA 02141, USA
All repeats were identified using RepeatMasker:
Smit, A.F.A. & Green, P. (1996-1997)
http://ftp.genome.washington.edu/RM/RepeatMasker.html
----- Genome Center
Center: Whitehead Institute/ MIT Center for Genome Research
Center code: WIBR
Web site: http://www-seq.wi.mit.edu
Contact: sequence_submission@genome.wi.mit.edu
----- Project Information
Center project name: L4861
Center clone name: 30_C_21
-----
* NOTE: This record contains 87 individual
* sequencing reads that have not been assembled into
* contigs. Runs of N are used to separate the reads
* and the order in which they appear is completely
* arbitrary. Low-pass sequence sampling is useful for
* identifying clones that may be gene-rich and allows
* overlap relationships among clones to be deduced.
* However, it should not be assumed that this clone
* will be sequenced to completion. In the event that
* the record is updated, the accession number will
* be preserved.
1 741: contig of 741 bp in length
gap of unknown length
742 1519: contig of 778 bp in length
gap of unknown length
1520 2316: contig of 797 bp in length
gap of unknown length
2317 3103: contig of 787 bp in length
gap of unknown length
3104 3870: contig of 767 bp in length
gap of unknown length
3871 4640: contig of 770 bp in length
gap of unknown length
4641 5433: contig of 793 bp in length
gap of unknown length
5434 6223: contig of 790 bp in length
gap of unknown length
6224 6976: contig of 753 bp in length
gap of unknown length
6977 7737: contig of 761 bp in length
gap of unknown length
7738 8503: contig of 766 bp in length
gap of unknown length
8504 9287: contig of 784 bp in length
gap of unknown length
9288 10066: contig of 779 bp in length
gap of unknown length
10067 10852: contig of 786 bp in length
gap of unknown length
10853 11621: contig of 769 bp in length
gap of unknown length
11622 12411: contig of 790 bp in length
gap of unknown length
12412 13191: contig of 780 bp in length
gap of unknown length
13192 13983: contig of 792 bp in length
gap of unknown length
13984 14772: contig of 789 bp in length
gap of unknown length
14773 15538: contig of 766 bp in length
gap of unknown length
15539 16299: contig of 761 bp in length
gap of unknown length
16300 17091: contig of 792 bp in length
gap of unknown length
17092 17829: contig of 738 bp in length
gap of unknown length
17830 18630: contig of 801 bp in length
gap of unknown length
18631 19415: contig of 785 bp in length
gap of unknown length
19416 20205: contig of 790 bp in length
gap of unknown length
20206 20965: contig of 760 bp in length
gap of unknown length
20966 21730: contig of 765 bp in length
gap of unknown length
21731 22511: contig of 781 bp in length
gap of unknown length
22512 23290: contig of 779 bp in length
gap of unknown length
23291 24068: contig of 778 bp in length
gap of unknown length
24069 24841: contig of 773 bp in length
gap of unknown length
24842 25616: contig of 775 bp in length
gap of unknown length
25617 26401: contig of 785 bp in length
gap of unknown length
26402 27160: contig of 759 bp in length
gap of unknown length
27161 27919: contig of 759 bp in length
gap of unknown length
27920 28716: contig of 797 bp in length
gap of unknown length
28717 29511: contig of 795 bp in length
gap of unknown length
29512 30270: contig of 759 bp in length
gap of unknown length
30271 31050: contig of 780 bp in length
gap of unknown length
31051 31819: contig of 769 bp in length
gap of unknown length
31820 32597: contig of 778 bp in length
gap of unknown length
32598 33379: contig of 782 bp in length
gap of unknown length
33380 34168: contig of 789 bp in length
gap of unknown length
34169 34930: contig of 762 bp in length
gap of unknown length

```



- \* NOTE: This is a 'working draft' sequence.
- \* This record will be updated with the finished sequence
- \* as soon as it is available and the accession number will
- \* be preserved.

```
Location/Qualifiers
1. .190839
/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="10"
/clone="p011-155N8"
```

Query Match 1.5%; Score 34; DB 23; Length 190839;  
Best Local Similarity 86.8%; Pred. NO. 3.42e-05;  
Matches 46; Conservative 0; Mismatches 6; Indels 1; Gaps 1;

[illegible]

ACCESSION AC015864

VERSION AC015864.1 GI:6446804

**KEYWORDS** HTG; HTGS\_PHASE0; NULL.

**SOURCE** human.

ORGANISM	Homo sapiens
1	1
2	2
3	3
4	4
5	5
6	6
7	7
8	8
9	9
10	10
11	11
12	12
13	13
14	14
15	15
16	16
17	17
18	18
19	19
20	20
21	21
22	22
23	23
24	24
25	25
26	26
27	27
28	28
29	29
30	30
31	31
32	32
33	33
34	34
35	35
36	36
37	37
38	38
39	39
40	40
41	41
42	42
43	43
44	44
45	45
46	46
47	47
48	48
49	49
50	50
51	51
52	52
53	53
54	54
55	55
56	56
57	57
58	58
59	59
60	60
61	61
62	62
63	63
64	64
65	65
66	66
67	67
68	68
69	69
70	70
71	71
72	72
73	73
74	74
75	75
76	76
77	77
78	78
79	79
80	80
81	81
82	82
83	83
84	84
85	85
86	86
87	87
88	88
89	89
90	90
91	91
92	92
93	93
94	94
95	95
96	96
97	97
98	98
99	99
100	100

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

REFERENCE  
1 (bases 1 to 1927/98)

**AUTHORS** Birrell, B., Linton, L., Nusbaum, C. and Lander, E.  
**TITLE** Homo sapiens chromosome 18 clone RP11-54A1

JOURNAL  
of  
HOMOSEXUALITY  
Unpublished

REFERENCE 2 (bases 1 to 192798)

Birren, B., Linton, L., Nusbaum, C., Lander, E., Allen, N., Anderson, M., Baldwin, J., Barna, N., Beckerly, R., Boguslavsky, L., Bouhgalter, B., Brown, A., Castle, A., Collangelo, M., Collins, S., Collymore, A., Cooke, P., DeAtellano, K., Dewar, K., Domino, M., Donelan, L., Doyle, M., Ferreira, P., FitzHugh, W., Forrest, C., Funke, R., Gage, D., Galagan, J., Gardyna, S., Grant, G., Hagos, B., Heaford, A., Horton, L., Howland, J. C., Johnson, R., Jones, C., Kann, L., Karatas, A., Klein, J.,

Lehoczyk, J., Lieu, C., Locke, K., Macdonald, P., Marquis, N., McEwan, P., McGurk, A., McKernan, K., McLaughlin, J., Meldrum, J., Morrow, J., Naylor, J., Norman, C.H., O'Connor, T., O'Donnell, P., Peterson, K., Pollara, V., Riley, R., Roy, A., Santos, R., Severy, P., Stange-Thomann, N., Stojanovic, N., Subramanian, A., Talamas, J., Tesfaye, S., Tirrell, A., Vassiliev, H., Vo, A., Wheeler, J., Wu, X., Wyman, D., Ye, W.J., Zimmer, A. and Zody, M.

Direct Submission  
Submitted (17-NOV-1999) Whitehead Institute/MIT Center for Genome Research, 320 Charles Street, Cambridge, MA 02141, USA  
All repeats were identified using RepeatMasker:  
Smit, A.F.A. & Green, P. (1996-1997)  
<http://ftp.genome.washington.edu/RM/RepeatMasker.html>

----- Project Information  
Center project name: L625  
Center clone name: 54\_A\_1  
-----

\* NOTE: This record contains 234 individual  
\* sequencing reads that have not been assembled into  
\* contigs. Runs of N are used to separate the reads  
\* and the order in which they appear is completely  
\* arbitrary. Low-pass sequence sampling is useful for  
\* identifying clones that may be gene-rich and allows  
\* overlap relationships among clones to be deduced.  
\* However, it should not be assumed that this clone  
\* will be sequenced to completion. In the event that  
\* the record is updated, the accession number will  
\* be preserved.

\* 1 775: contig of 775 bp in length  
\* gap of unknown length  
\* 776 contig of 758 bp in length  
\* gap of unknown length  
\* 1534 2303: contig of 770 bp in length  
\* gap of unknown length  
\* 2304 3054: contig of 751 bp in length  
\* gap of unknown length  
\* 3055 3833: contig of 779 bp in length  
\* gap of unknown length  
\* 3834 4596: contig of 763 bp in length  
\* gap of unknown length  
\* 4597 5342: contig of 746 bp in length  
\* gap of unknown length  
\* 5343 6084: contig of 742 bp in length  
\* gap of unknown length  
\* 6085 6842: contig of 758 bp in length  
\* gap of unknown length  
\* 6843 7590: contig of 748 bp in length  
\* gap of unknown length  
\* 7591 8347: contig of 757 bp in length  
\* gap of unknown length  
\* 8348 9100: contig of 753 bp in length  
\* gap of unknown length  
\* 9101 9878: contig of 778 bp in length  
\* gap of unknown length  
\* 9879 10638: contig of 760 bp in length  
\* gap of unknown length  
\* 10639 11405: contig of 767 bp in length  
\* gap of unknown length  
\* 11406 12216: contig of 811 bp in length  
\* gap of unknown length  
\* 12217 12986: contig of 770 bp in length  
\* gap of unknown length  
\* 12987 13746: contig of 760 bp in length  
\* gap of unknown length  
\* 13747 14520: contig of 774 bp in length  
\* gap of unknown length  
\* 14521 15256: contig of 736 bp in length  
\* gap of unknown length  
\* 15257 16019: contig of 763 bp in length  
\* gap of unknown length  
\* 16020 16781: contig of 762 bp in length  
\* gap of unknown length  
\* 16782 17512: contig of 731 bp in length

\* 17513 18283: contig of 771 bp in length  
\* gap of unknown length  
\* 18284 19039: contig of 756 bp in length  
\* gap of unknown length  
\* 19040 19814: contig of 775 bp in length  
\* gap of unknown length  
\* 19815 20576: contig of 762 bp in length  
\* gap of unknown length  
\* 20577 21234: contig of 658 bp in length  
\* gap of unknown length  
\* 21235 21992: contig of 758 bp in length  
\* gap of unknown length  
\* 21993 22737: contig of 745 bp in length  
\* gap of unknown length  
\* 22738 23510: contig of 773 bp in length  
\* gap of unknown length  
\* 23511 24255: contig of 745 bp in length  
\* gap of unknown length  
\* 24256 25060: contig of 805 bp in length  
\* gap of unknown length  
\* 25061 25822: contig of 762 bp in length  
\* gap of unknown length  
\* 25823 26682: contig of 860 bp in length  
\* gap of unknown length  
\* 26683 27437: contig of 755 bp in length  
\* gap of unknown length  
\* 27438 28187: contig of 750 bp in length  
\* gap of unknown length  
\* 28188 28936: contig of 749 bp in length  
\* gap of unknown length  
\* 28937 29707: contig of 771 bp in length  
\* gap of unknown length  
\* 29708 30473: contig of 766 bp in length  
\* gap of unknown length  
\* 30474 31233: contig of 760 bp in length  
\* gap of unknown length  
\* 31234 31980: contig of 747 bp in length  
\* gap of unknown length  
\* 31981 32748: contig of 768 bp in length  
\* gap of unknown length  
\* 32749 33519: contig of 771 bp in length  
\* gap of unknown length  
\* 33520 34282: contig of 763 bp in length  
\* gap of unknown length  
\* 34283 35049: contig of 767 bp in length  
\* gap of unknown length  
\* 35050 35822: contig of 773 bp in length  
\* gap of unknown length  
\* 35823 36597: contig of 775 bp in length  
\* gap of unknown length  
\* 36598 37346: contig of 749 bp in length  
\* gap of unknown length  
\* 37347 38099: contig of 753 bp in length  
\* gap of unknown length  
\* 38100 38799: contig of 700 bp in length  
\* gap of unknown length  
\* 38800 39542: contig of 743 bp in length  
\* gap of unknown length  
\* 39543 40306: contig of 764 bp in length  
\* gap of unknown length  
\* 40307 41058: contig of 752 bp in length  
\* gap of unknown length  
\* 41059 41821: contig of 763 bp in length  
\* gap of unknown length  
\* 41822 42562: contig of 741 bp in length  
\* gap of unknown length  
\* 42563 43302: contig of 740 bp in length  
\* gap of unknown length  
\* 43303 44064: contig of 762 bp in length  
\* gap of unknown length  
\* 44065 44857: contig of 793 bp in length  
\* gap of unknown length

TITLE  
JOURNAL  
COMMENT

```

* 44858 45617: contig of 760 bp in length
* gap of unknown length
* 45618 46385: contig of 768 bp in length
* gap of unknown length
* 46386 47154: contig of 769 bp in length
* gap of unknown length
* 47155 47915: contig of 761 bp in length
* gap of unknown length
* 47916 48659: contig of 744 bp in length
* gap of unknown length
* 48660 49442: contig of 783 bp in length
* gap of unknown length
* 49443 50204: contig of 762 bp in length
* gap of unknown length
* 50205 50981: contig of 777 bp in length
* gap of unknown length
* 50982 51746: contig of 765 bp in length
* gap of unknown length
* 51747 52515: contig of 769 bp in length
* gap of unknown length
* 52516 53383: contig of 868 bp in length
* gap of unknown length
* 53384 54171: contig of 788 bp in length
* gap of unknown length
* 54172 54939: contig of 768 bp in length
* gap of unknown length
* 54940 55702: contig of 763 bp in length
* gap of unknown length
* 55703 56477: contig of 775 bp in length
* gap of unknown length
* 56478 57239: contig of 762 bp in length
* gap of unknown length

```

....  
Note: remainder of annotations omitted.

```

Query Match      1.5%; Score 34; DB 26; Length 192798;
Best Local Similarity 76.6%; Pred. No. 3.42e-05;
Matches 49; Conservative 0; Mismatches 15; Indels 0; Gaps 0;

Db 124736 TTTAATAATATGTTGTCATGACGCAAAATATACAGAAATATTATTTCAACATGTGATCAAT 124795
|||||
Qy 2226 TTTAATAATATATCTTATGAGCCCTATATATTCAAAATATTATGTTAACAATGTAATCCAT 2285
|||||

Db 124796 ATTT 124799
|||
Qy 2286 GTTT 2289

```

Search completed: Thu Apr 20 16:13:59 2000  
Job time : 11804 secs.



\*\*\*\*\*

WIREH

(TM)

\*\*\*\*\*

Release 3.1A John F. Collins, Biocomputing Research Unit.  
Copyright (c) 1993-1998 University of Edinburgh, U.K.  
Distribution rights by Oxford Molecular Ltd

MPsrch\_un n.a. - n.a. database search, using Smith-Waterman algorithm

Run on: Sun Apr 23 09:36:33 2000; MasPar time 3972.39 Seconds  
Tabular output not generated: 1298.715 Million cell updates/sec

Title: >US-09-249-003-1  
Description: (1-2297) from US09249003.seq  
Perfect Score: 2297  
N.A. Sequence: 1 CGGCTGTCTGGCAGTCTT.....TAATCATGTTCTTTTCC 2297  
Comp: GCCGACACACGCCGACAGAA.....ATTAGTACAAAGAAAGG

Scoring table: TABLE default  
Gap 6

Nmatch STD : Dbase 0; Query 0

Searched: 623196 seqs, 1122988046 bases x 2

Post-processing: Minimum Match 0%  
Listing first 45 summaries

Database: genbank2-115  
1:gb\_pat 2:gb\_ph 3:gb\_pl1 4:gb\_pl2 5:gb\_pl3 6:gb\_pr1  
7:gb\_pr2 8:gb\_pr3 9:gb\_pr4 10:gb\_ro 11:gb\_sts 12:gb\_sy  
13:gb\_un 14:gb\_vi

Statistics: Mean 11.531; Variance 5.769; scale 1.999

Pred. No. is the number of results predicted by chance to have a  
score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match %	Length	ID	Description	Pred. No.
1	2297	100.0	2297	6 HUMIDSX	Human iduronate 2-sulf	0.00e+00
2	2297	100.0	2297	1 I92327	Sequence 1 from patent	0.00e+00
3	2297	100.0	2297	1 AR025358	Sequence 1 from patent	0.00e+00
4	1113	48.5	1314	7 HUMIDSALT	Homo sapiens iduronate	0.00e+00
5	1096	47.7	1750	10 MUSIDS	Mus musculus iduronate	0.00e+00
6	996	43.4	1120	6 HUMIDSGEN9	Homo sapiens iduronate	0.00e+00
7	996	43.4	4428	1 AR025360	Sequence 6 from patent	0.00e+00
8	996	43.4	4428	1 I92329	Sequence 6 from patent	0.00e+00
9	995	43.3	36845	6 HUMIDS	Homo sapiens iduronate	0.00e+00
c 10	995	43.3	326663	9 AF011889	Human Xq28 cosmid U12	0.00e+00
11	437	19.0	1098	8 AF050145	Homo sapiens iduronate	0.00e+00
12	434	18.9	1831	10 MMRNA15	M.musculus mRNA for id	0.00e+00
13	390	17.0	400	11 G13552	human STS SHGC-11149.	1.49e-248
14	340	14.8	110079	10 AC002315	Mouse BAC-146N21 Chrom	1.86e-248
c 15	231	10.1	233	1 AC002315	Sequence 929 from Pate	1.60e-157
c 16	231	10.1	233	1 A75243	Sequence 929 from Pate	1.60e-157
17	229	10.0	536	6 HUMIDSGEN1	Homo sapiens iduronate	7.22e-156
18	205	8.9	322	6 HUMISS05	Homo sapiens iduronate	4.33e-136

19 205 8.9 490 6 HUMIDSGEN5 Homo sapiens iduronate 4.33e-136  
c 20 186 8.1 297 7 H5U66055 Human clone PI5a-1 IDS 1.59e-120  
21 186 8.1 425 7 H5U66054 Human clone PI3b-2 IDS 1.59e-120  
c 22 187 8.1 487 7 H5U66053 Human clone PI3b-1 IDS 2.43e-121  
23 182 7.9 298 6 HUMISS03 Homo sapiens iduronate 2.94e-117  
24 182 7.9 388 6 HUMIDSGEN3 Homo sapiens iduronate 2.94e-117  
c 25 182 7.9 627 8 AF050146 Homo sapiens clone 488 2.94e-117  
26 182 7.9 750 7 HUMIDS02 Homo sapiens iduronate 2.94e-117  
27 182 7.9 326663 9 AF011889 Human Xq28 cosmid U12 2.94e-117  
28 178 7.7 290 6 HUMISS06 Homo sapiens iduronate 5.35e-114  
29 176 7.7 294 6 HUMISS08 Homo sapiens iduronate 2.27e-112  
30 178 7.7 379 6 HUMIDSGEN6 Homo sapiens iduronate 5.35e-114  
31 176 7.7 404 6 HUMIDSGEN8 Homo sapiens iduronate 2.27e-112  
32 176 7.7 2331 8 HUMIDS71 Human (clone pBI7-8) i 2.27e-112  
33 139 6.1 257 6 HUMISS02 Homo sapiens iduronate 1.68e-82  
34 139 6.1 426 6 HUMIDSGEN2 Homo sapiens iduronate 1.68e-82  
35 132 5.7 248 6 HUMISS07 Homo sapiens iduronate 6.46e-77  
36 132 5.7 319 6 HUMIDSGEN7 Homo sapiens iduronate 6.46e-77  
37 105 4.6 162 6 HUMISS01 Homo sapiens iduronate 1.17e-55  
38 91 4.0 366 6 HUMIDSGEN4 Homo sapiens iduronate 7.27e-45  
39 90 3.9 169 6 HUMISS04 Homo sapiens iduronate 4.21e-44  
40 65 2.8 326 7 HUMIDS01 Homo sapiens iduronate 1.72e-25  
c 42 47 2.0 215 1 I28278 Sequence 14 from patent 2.46e-23  
43 45 2.0 74371 8 AC005369 Homo sapiens chromosom 6.19e-13  
c 44 40 1.7 131389 9 AC005728 Homo sapiens Chromosom 1.31e-11  
c 45 36 1.6 965 1 AR024229 Homo sapiens Chromosom 2.23e-08  
Sequence 22 from patent 6.74e-06

ALIGNMENTS

RESULT 1 HUMIDSX 2297 bp mRNA PRI 07-MAR-1995  
LOCUS Human iduronate 2-sulfatase mRNA, complete cds.  
DEFINITION M58342 M38371  
ACCESSION M58342.1 GI:184561  
VERSION iduronate-2-sulfatase.  
KEYWORDS Human, cDNA to mRNA.  
SOURCE Homo sapiens  
ORGANISM Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;  
Eutheria; Primates; Catarrhini; Hominiidae; Homo.  
REFERENCE 1 (bases 1 to 2297)  
AUTHORS Willson,P.J., Morris,C.P., Anson,D.S., Occhiodoro,T., Bielicki,J.,  
Clements,P.R. and Hopwood,J.J.  
TITLE Hunter syndrome: Isolation of an iduronate-2-sulfatase cDNA clone  
and analysis of patient DNA.  
JOURNAL Proc. Natl. Acad. Sci. U.S.A. 87 (21), 8531-8535 (1990)  
MEDLINE 91046030  
FEATURES Location/Qualifiers  
source 1. .2297  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/clone="lamda c2s23"  
/cell\_type="endothelial"  
/map="Xq27.3-q28"  
125. .199  
/gene="IDS"  
/note="G00-120-521"  
125. .1777  
/gene="IDS"  
/EC\_number="3.1.6.13"  
/db\_xref="GDB:G00-120-521"  
/db\_xref="GDB:G00-120-521"  
/product="iduronate 2-sulfatase"  
/protein\_id="AA63197.1"  
/db\_xref="GI:184562"  
/translation="MPPPTGRGLLGLVLSVVCVALGSETQANSTTDALNVLLIIV  
DLPRLSGVGYGDKLVRSNPIDQLASHLLFQNAFAQAVCAVRSVLTGRPPRTTL  
YDFNSVYRVHAGNFSTIPQYKENGYYVTVSGVKGFGHGISNHTDDSPYGSWSPYPH  
SSPKYENTKCRGPDGELHANLLCPVDLVDVPEGLPDKGQEQAIQLLEKMTSASP  
FFLAVGHKHPHPRPKFKQKLYPLENITLADPEVDPGLPPVAYNPWMDIRQEDY  
QALNISVPYGPIDPQFKIRQSFASVLDQVGRLLSLDLDQANSTIAFTSD  
HGWALGEHGEWAKYSNFDVATHVPLIFYVPGRTASLPAGEKLPFYLPDPDSASQLME

sig\_peptide

CDS

PGROSDVLVSLFPTLAGLAGLOVPRCPVPSFHVFLCBEGKNLLKHFRPRDLEED  
PYLPNCNRELIAYSQYPRPSDIPQWNSDKPSLKDKIKMGYSIRITIDRYTIVWVGFNPD  
EFLANFSDIIFAGELYFVDSODPLQDHNMYNSQGGDLFQLLMP\*

gene

/gene="IDS"

200..1777

mat\_peptide

/note="G00-120-521"

BASE COUNT 553 a 600 c 510 g 634 t

ORIGIN

Query Match		100.0%;	Score 2297;	DB 6;	Length 2297;
Best Local Similarity		100.0%;	Pred. No. 0.00e+00;		
Matches 2297;		Conservative	0;	Mismatches	0;
				Indels	0;
				Gaps	0;
Db	1	CGGCTGTGTTGGCAGCTCTTCATGGTTCGGACAGAGAGGTCTCTGTGGCTGCGCGGCG 60			
QY	1	CGGCTGTGTTGGCAGCTCTTCATGGTTCGGACAGAGAGGTCTCTGTGGCTGCGCGGCG 60			
Db	61	TGCTAACTGCGCCACCTGCTGAGCCTGTCCCGCGCTCTGAAAGCGCGCGTCAAGC 120			
QY	61	TGCTAACTGCGCCACCTGCTGAGCCTGTCCCGCGCTCTGAAAGCGCGCGTCAAGC 120			
Db	121	CGAAATGCGCCACCCCGAGCGGCGGAGGCTTCTCTGGCTGGGTCTGGTCTGAGCTC 180			
QY	121	CGAAATGCGCCACCCCGAGCGGCGGAGGCTTCTCTGGCTGGGTCTGGTCTGAGCTC 180			
Db	181	CGTCTGCTGCCCTCGGATCGGAAACGAGGCCAACTCGACACAGATGCTCTGAAGCT 240			
QY	181	CGTCTGCTGCCCTCGGATCGGAAACGAGGCCAACTCGACACAGATGCTCTGAAGCT 240			
Db	241	TCCTTCTCATATCGTGGATGACCTGCGGCCCTCCCTGGGCTGTTATGGGGATAAGCTGT 300			
QY	241	TCCTTCTCATATCGTGGATGACCTGCGGCCCTCCCTGGGCTGTTATGGGGATAAGCTGT 300			
Db	301	GAGTCCCAATATTGACCAACTGGGATCCGACAGCCCTCTCTCCAGAAATGCGCTTGC 360			
QY	301	GAGTCCCAATATTGACCAACTGGGATCCGACAGCCCTCTCTCCAGAAATGCGCTTGC 360			
Db	361	GCAGCAAGCAGTGTGCGCCCGAGCGCGTTCCTTCTCTCACTGCGAGGAGACCTGACAC 420			
QY	361	GCAGCAAGCAGTGTGCGCCCGAGCGCGTTCCTTCTCTCACTGCGAGGAGACCTGACAC 420			
Db	421	CACCCGCTGTACGACTTCAACTCTACTGGAGGCTGCAGCGTGGAAACTTCTCCACCAT 480			
QY	421	CACCCGCTGTACGACTTCAACTCTACTGGAGGCTGCAGCGTGGAAACTTCTCCACCAT 480			
Db	481	CCCCAGTACTTCAAGGAGATGGCTATGTGACCATGTGCGTGGGAAAGTCTTTCAACC 540			
QY	481	CCCCAGTACTTCAAGGAGATGGCTATGTGACCATGTGCGTGGGAAAGTCTTTCAACC 540			
Db	541	TGGGATATCTTCTAACCATACCGATGATTCGCCGTATAGCTGGTCTTTCCACCTATCA 600			
QY	541	TGGGATATCTTCTAACCATACCGATGATTCGCCGTATAGCTGGTCTTTCCACCTATCA 600			
Db	601	TCCTTCTCTCAGAGATGAAACACTAAGACATGTGAGGGCCAGATGGAACTCCA 660			
QY	601	TCCTTCTCTCAGAGATGAAACACTAAGACATGTGAGGGCCAGATGGAACTCCA 660			
Db	661	TGCCAACCTGCTTGGCCCTGTGGATGTGCTGGATGTTCCCGAGGCGACTTGCCTGACAA 720			
QY	661	TGCCAACCTGCTTGGCCCTGTGGATGTGCTGGATGTTCCCGAGGCGACTTGCCTGACAA 720			
Db	721	ACAGAGACTGAGCAAGCCATACAGTGTGGAAAAGATGAAAAGTCAAGCAGTCCCTTT 780			
QY	721	ACAGAGACTGAGCAAGCCATACAGTGTGGAAAAGATGAAAAGTCAAGCAGTCCCTTT 780			
Db	781	CTTCTGCGGCTTGGGTATACAGCCACATCCCTTCAGATACCCCAAGGAATTTCA 840			
QY	781	CTTCTGCGGCTTGGGTATACAGCCACATCCCTTCAGATACCCCAAGGAATTTCA 840			
Db	841	GAACTGTATCCCTTGGAGAACATCACCTTGCGCCCGCATCCCGAGGTCCCTTGATGGCT 900			
QY	841	GAACTGTATCCCTTGGAGAACATCACCTTGCGCCCGCATCCCGAGGTCCCTTGATGGCT 900			

QY	841	GAAGTTGTATCCCTTGGAGAACATCACCTTGGCCCGCGATCCGAGGTCCCTGATGGCT 900	
Db	901	ACCCCTGTGGCTTACAAACCCCTGGATGGACATCAGCAACGSGAAGAGCTCCAAAGCTT 960	
QY	901	ACCCCTGTGGCTTACAAACCCCTGGATGGACATCAGCAACGSGAAGAGCTCCAAAGCTT 960	
Db	961	AAACATCAGTGTGCGGTATGGTCCAAATTCCTGTGGACTTTTCAGCGGAAAATCCGCGAG 1020	
QY	961	AAACATCAGTGTGCGGTATGGTCCAAATTCCTGTGGACTTTTCAGCGGAAAATCCGCGAG 1020	
Db	1021	CTACTTTGGCTCTGTGTCATATTTGGATACACAGGTGCGCGCTCTTGAGTCTTTTGA 1080	
QY	1021	CTACTTTGGCTCTGTGTCATATTTGGATACACAGGTGCGCGCTCTTGAGTCTTTTGA 1080	
Db	1081	CGATCTTCAGCTGGCCACAGACACCATCATTTGATTTTACCTCGGATCAGTGGTGGCTCT 1140	
QY	1081	CGATCTTCAGCTGGCCACAGACACCATCATTTGATTTTACCTCGGATCAGTGGTGGCTCT 1140	
Db	1141	AGGTGAACATGGAGAAATGGCCAAATACAGCAATTTGATTTGCTACCCATGTTCCCT 1200	
QY	1141	AGGTGAACATGGAGAAATGGCCAAATACAGCAATTTGATTTGCTACCCATGTTCCCT 1200	
Db	1201	GATATTCATGTTCTCGGAAGGACGCTTCACCTCCGAGGAGCGGAGAGCTTTTCCC 1260	
QY	1201	GATATTCATGTTCTCGGAAGGACGCTTCACCTCCGAGGAGCGGAGAGCTTTTCCC 1260	
Db	1261	TTACTTCGACCCCTTTTGGATTTCCGCTCACAGTTGATGGAGCAGGAGCAATCCATGA 1320	
QY	1261	TTACTTCGACCCCTTTTGGATTTCCGCTCACAGTTGATGGAGCAGGAGCAATCCATGA 1320	
Db	1321	CCCTTGTGGAACCTGTGTCTCTTTTCCACGCTGGCTGGACTTGCAGGACTTCAGGTTCC 1380	
QY	1321	CCCTTGTGGAACCTGTGTCTCTTTTCCACGCTGGCTGGACTTGCAGGACTTCAGGTTCC 1380	
Db	1381	ACTCGCTCCCGCTTCCCTTCATTTTCCAGCTGAGCTGTCGAGAGGAGCAAGCACTTCT 1440	
QY	1381	ACTCGCTCCCGCTTCCCTTCATTTTCCAGCTGAGCTGTCGAGAGGAGCAAGCACTTCT 1440	
Db	1441	GAAGCAATTTTCGATTCGCTGACTTGGAGAGGATCCGTACCTCCCTGGAATCCCGTGA 1500	
QY	1441	GAAGCAATTTTCGATTCGCTGACTTGGAGAGGATCCGTACCTCCCTGGAATCCCGTGA 1500	
Db	1501	ACTGATTTGCTATAGCAGTATCCCGGCTTTCAGACATCCCTCAGTGGAAATTCGACAA 1560	
QY	1501	ACTGATTTGCTATAGCAGTATCCCGGCTTTCAGACATCCCTCAGTGGAAATTCGACAA 1560	
Db	1561	GCCGAGTTTAAAGATATAAAGATCATGGCTATTCATACGCAACATAGACTATAGGTA 1620	
QY	1561	GCCGAGTTTAAAGATATAAAGATCATGGCTATTCATACGCAACATAGACTATAGGTA 1620	
Db	1621	TACTGTGGTGGTGGCTTCAATCCTGATGAATTTCTAGCTAACTTTCTGACATCCATGC 1680	
QY	1621	TACTGTGGTGGTGGCTTCAATCCTGATGAATTTCTAGCTAACTTTCTGACATCCATGC 1680	
Db	1681	AGGGAACCTGATTTTGTGGATTCGACCATTCAGGATCACAATATGATTAATGATTC 1740	
QY	1681	AGGGAACCTGATTTTGTGGATTCGACCATTCAGGATCACAATATGATTAATGATTC 1740	
Db	1741	CCAAGTGGAGATCTTTTCCAGTTGTGCTTTCAGTTTTCGCAACCATGGATGGCAA 1800	
QY	1741	CCAAGTGGAGATCTTTTCCAGTTGTGCTTTCAGTTTTCGCAACCATGGATGGCAA 1800	
Db	1801	ATGTGATGTGCTCCCTTCCAGCTGGTGGAGGAGGAGTTAGAGCTGGTGGTGGTGA 1860	
QY	1801	ATGTGATGTGCTCCCTTCCAGCTGGTGGAGGAGGAGTTAGAGCTGGTGGTGGTGA 1860	
Db	1861	ACCCATAATATTGGAAGAGCCTGAGGCTAGTTTATCCAAACATGATCAATTTGG 1920	
QY	1861	ACCCATAATATTGGAAGAGCCTGAGGCTAGTTTATCCAAACATGATCAATTTGG 1920	
Db	1921	CCTGAGAATATCTAACAGCAACCTTTTCTGTTTCTGCTTTTAAATTTTAAATGCT 1980	
QY	1921	CCTGAGAATATCTAACAGCAACCTTTTCTGTTTCTGCTTTTAAATTTTAAATGCT 1980	

Db 1981 AATTGGACAGTTTTTTTTTAAATTTCCCTCTTTTAAACAGTTACGGCTTATTACTG 2040  
QY 1981 AATTGGACAGTTTTTTTTTAAATTTCCCTCTTTTAAACAGTTACGGCTTATTACTG 2040  
Db 2041 AATAAATACAAAGCAAACTCAAGTTATGTCATACCTTTGGATACGAAGACCATACA 2100  
QY 2041 AATAAATACAAAGCAAACTCAAGTTATGTCATACCTTTGGATACGAAGACCATACA 2100  
Db 2101 TAATAACCAACATACATATATACAAAGAAATACATTTTCATATTTTGTGAAATTTAGTGC 2160  
QY 2101 TAATAACCAACATACATATATACAAAGAAATACATTTTCATATTTTGTGAAATTTAGTGC 2160  
Db 2161 ATTTCAAAAAGTAATCATATATATCAAACTAGGCACCACTAAGTTCCTGATTTATTTGTT 2220  
QY 2161 ATTTCAAAAAGTAATCATATATATCAAACTAGGCACCACTAAGTTCCTGATTTATTTGTT 2220  
Db 2221 TATAATTTAATATATATCTTATGAGCCCTATATATTTCAAAATATTATGTTAACAATGTA 2280  
QY 2221 TATAATTTAATATATCTTATGAGCCCTATATATTTCAAAATATTATGTTAACAATGTA 2280  
Db 2281 TCCATGTTCTTTTCC 2297  
QY 2281 TCCATGTTCTTTTCC 2297

RESULT 2  
LOCUS 192327 2297 bp DNA PAT 01-DEC-1998  
DEFINITION Sequence 1 from patent US 5728381.  
ACCESSION 192327  
VERSION 192327.1 GI:3936797  
KEYWORDS  
SOURCE Unknown.  
ORGANISM Unknown.  
REFERENCE 1 (bases 1 to 2297)  
AUTHORS Wilson,P.J., Morris,C.Phillip, Anson,D.Stewart, Occhiodoro,T.,  
Bielicki,J., Clements,P.Roy and Hopwood,J.Joseph.  
TITLE Glycosylation variants of Iduronate 2-sulfatase  
JOURNAL Patent: US 5728381-A 1 17-MAR-1998;  
FEATURES  
source 1. 2297  
BASE COUNT 553 a 600 c 510 g 634 t  
ORIGIN

Query Match 100.0%; Score 2297; DB 1; Length 2297;  
Best Local Similarity 100.0%; Pred.No. 0.00e+00;  
Matches 2297; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Db 1 CGGCTGTGTGCGCAGTCTTCATGGTTCCCGACGAGGAGTCTCTGTGGCTCGCGCGGC 60  
QY 1 CGGCTGTGTGCGCAGTCTTCATGGTTCCCGACGAGGAGTCTCTGTGGCTCGCGCGGC 60  
Db 61 TGTCTACTCGCGCAGTCTGTGAGCCTGTCCCGCGCTCTGAAGCGCGCGCTCGAAGC 120  
QY 61 TGTCTACTCGCGCAGTCTGTGAGCCTGTCCCGCGCTCTGAAGCGCGCGCTCGAAGC 120  
Db 121 CGAAATGCGCGCCACCGCGCGGAGGCTTCTCTGGCTGGTCTGGTCTGAGCTC 180  
QY 121 CGAAATGCGCGCCACCGCGCGGAGGCTTCTCTGGCTGGTCTGGTCTGAGCTC 180  
Db 181 CGTCTCGCTCGCGCTCGGATCGAAACGAGGCAACTCGACCACAGATGCTCTGAACGT 240  
QY 181 CGTCTCGCTCGCGCTCGGATCGAAACGAGGCAACTCGACCACAGATGCTCTGAACGT 240  
Db 241 TCTTCTCATCATCTGGATGACTCGCGCCCTCCCTGGGCTGTATGGGATGAAGCTGGT 300  
QY 241 TCTTCTCATCATCTGGATGACTCGCGCCCTCCCTGGGCTGTATGGGATGAAGCTGGT 300  
Db 301 GAGGTCCCAATATTGACCAACTGCGATCCGACAGCCCTCTTCAGAAATGCTTTGC 360  
QY 301 GAGGTCCCAATATTGACCAACTGCGATCCGACAGCCCTCTTCAGAAATGCTTTGC 360

Db 361 GCAGCAAGCAGTGTGCGCCCGAGCGGTTTCTTCTCTCACTGCAGGAGACCTGACAC 420  
QY 361 GCAGCAAGCAGTGTGCGCCCGAGCGGTTTCTTCTCTCACTGCAGGAGACCTGACAC 420  
Db 421 CACCCGCTGTACGACTTCAACTCCTACTGGAGGTGCACGCTGGAACATTTCTCCACAT 480  
QY 421 CACCCGCTGTACGACTTCAACTCCTACTGGAGGTGCACGCTGGAACATTTCTCCACAT 480  
Db 481 CCCCCAGTACTTCAAGGAGAAATGCTATGACCATGTGCGTGGGAAAAGTCTTTTCACCC 540  
QY 481 CCCCCAGTACTTCAAGGAGAAATGCTATGACCATGTGCGTGGGAAAAGTCTTTTCACCC 540  
Db 541 TGGGATATCTTCTAACCACATACCGATGATCTCCGTATAGCTGGTCTTTTCCACCTTATCA 600  
QY 541 TGGGATATCTTCTAACCACATACCGATGATCTCCGTATAGCTGGTCTTTTCCACCTTATCA 600  
Db 601 TCCTTCTCTGAGAAATGATGAAACACTAAGACATGTGAGGGCCAGATGGAGAACTCCA 660  
QY 601 TCCTTCTCTGAGAAATGATGAAACACTAAGACATGTGAGGGCCAGATGGAGAACTCCA 660  
Db 661 TGGCAACCTGCTTTGCGCTGTGGATGTGTCGATGTTCCCGAGGGCACCTTGCCTGACAA 720  
QY 661 TGGCAACCTGCTTTGCGCTGTGGATGTGTCGATGTTCCCGAGGGCACCTTGCCTGACAA 720  
Db 721 ACAGAGCACTGAGCAAGCCATACAGTTGTTGGAAGATGAAACGTCAGCAGTCTCTT 780  
QY 721 ACAGAGCACTGAGCAAGCCATACAGTTGTTGGAAGATGAAACGTCAGCAGTCTCTT 780  
Db 781 CTTCCTGGCGTGTGGTATCATAGCCACACATCCCTTTCAGATACCCCAAGAAATTTCA 840  
QY 781 CTTCCTGGCGTGTGGTATCATAGCCACACATCCCTTTCAGATACCCCAAGAAATTTCA 840  
Db 841 GAAGTTGTATCCCTTGGAGAAACATCACCTGCGCCCGCATCCCGAGGTCCCTGATGGCT 900  
QY 841 GAAGTTGTATCCCTTGGAGAAACATCACCTGCGCCCGCATCCCGAGGTCCCTGATGGCT 900  
Db 901 ACCCCCTGTGGCTTACAAACCTCGATGGACATCAGGCAACGGGAAGAGCTTCAAGCCTT 960  
QY 901 ACCCCCTGTGGCTTACAAACCTCGATGGACATCAGGCAACGGGAAGAGCTTCAAGCCTT 960  
Db 961 AAACATCAGTGTGCCGTATGTCCTCAATTCCTGTGACATTTCCAGCGGAAATCCGCCAG 1020  
QY 961 AAACATCAGTGTGCCGTATGTCCTCAATTCCTGTGACATTTCCAGCGGAAATCCGCCAG 1020  
Db 1021 CTACTTTGCGCTCTGTGTATATTTGGATACACAGTGTGGCGCGCTCTTGGTGTCTTGA 1080  
QY 1021 CTACTTTGCGCTCTGTGTATATTTGGATACACAGTGTGGCGCGCTCTTGGTGTCTTGA 1080  
Db 1081 CGATCTTTCAGTGTGCCCAACAGCACCATCATTTGATTTTACCTCGGATCATGTTGGTCT 1140  
QY 1081 CGATCTTTCAGTGTGCCCAACAGCACCATCATTTGATTTTACCTCGGATCATGTTGGTCT 1140  
Db 1141 AGTGAACATGGAGAAATGGGCCAAATACAGCAATTTTGTATGTTGTACCCATGTTCCCT 1200  
QY 1141 AGTGAACATGGAGAAATGGGCCAAATACAGCAATTTTGTATGTTGTACCCATGTTCCCT 1200  
Db 1201 GATATTTCTATGTTCTGSAAGCGGTCTCACTTCCGAGGCGAGAGAGCTTTTCCC 1260  
QY 1201 GATATTTCTATGTTCTGSAAGCGGTCTCACTTCCGAGGCGAGAGAGCTTTTCCC 1260  
Db 1261 TTACTTCGACCTTTTGTATTCGCTCACAGTTGATGGAGCCAGCAGGCAATCCATGGA 1320  
QY 1261 TTACTTCGACCTTTTGTATTCGCTCACAGTTGATGGAGCCAGCAGGCAATCCATGGA 1320  
Db 1321 CTTTGTGGAATCTGTCTCTTTTCCACGCTGGCTGGACTTGCAGGACTTGCAGGTTC 1380  
QY 1321 CTTTGTGGAATCTGTCTCTTTTCCACGCTGGCTGGACTTGCAGGACTTGCAGGTTC 1380  
Db 1381 ACCTCGCTGCCCTTCTCTTTCATTTACGTTGAGCTGTGCAGAGAGCAAGACCTTCT 1440  
QY 1381 ACCTCGCTGCCCTTCTCTTTCATTTACGTTGAGCTGTGCAGAGAGCAAGACCTTCT 1440

Db 1441 GAAGCATTTTCGATTCCTGCTTGGAGAGGATCCGTACCTCCCTGGTAAATCCCGTGA 1500  
QY 1441 GAAGCATTTTCGATTCCTGCTTGGAGAGGATCCGTACCTCCCTGGTAAATCCCGTGA 1500  
Db 1501 ACTGATTCCTATAGCAGTATCCCGGCTTCAGACATCCCTCAGTGGAAATCTGACAA 1560  
QY 1501 ACTGATTCCTATAGCAGTATCCCGGCTTCAGACATCCCTCAGTGGAAATCTGACAA 1560  
Db 1561 GCCGAGTTTAAAGATATAAAGATCATGGGCTATTCATACGCCACCATAGACTATAGSTA 1620  
QY 1561 GCCGAGTTTAAAGATATAAAGATCATGGGCTATTCATACGCCACCATAGACTATAGSTA 1620  
Db 1621 TACTGTGTGGGTGGCTTCATTCCTGATGAATTTCTAGTAACTTTCTGACATCCATGC 1680  
QY 1621 TACTGTGTGGGTGGCTTCATTCCTGATGAATTTCTAGTAACTTTCTGACATCCATGC 1680  
Db 1681 AGGGAACTGTATTTGTGGATTCGACCCATTCGAGGATCAATATGATATAATGATTC 1740  
QY 1681 AGGGAACTGTATTTGTGGATTCGACCCATTCGAGGATCAATATGATATAATGATTC 1740  
Db 1741 CCAAGTGGAGATCTTTTCCAGTTGTTGATGCTTTGAGTCTTTTCCAAACCATGGATGGCAA 1800  
QY 1741 CCAAGTGGAGATCTTTTCCAGTTGTTGATGCTTTTCCAAACCATGGATGGCAA 1800  
Db 1801 ATGTGATGTCTCCCTTCCAGTGTGAGAGGAGGATTTAGAGCTGTGCTTTGTGATT 1860  
QY 1801 ATGTGATGTCTCCCTTCCAGTGTGAGAGGAGGATTTAGAGCTGTGCTTTGTGATT 1860  
Db 1861 ACCATAATATTGGAAGCAGCTGAGGCTAGTTAATCCAAATGCATCAACAATTTGG 1920  
QY 1861 ACCATAATATTGGAAGCAGCTGAGGCTAGTTAATCCAAATGCATCAACAATTTGG 1920  
Db 1921 CCTGAGATATGTAAACAGCCAAACCTTTTCTGTTAGTCTTTTAAATTTTAAATTTGTT 1980  
QY 1921 CCTGAGATATGTAAACAGCCAAACCTTTTCTGTTAGTCTTTTAAATTTTAAATTTGTT 1980  
Db 1981 AATTGGACCACTTTTAAATTTTCCCTCTTTTAAACAGTTTACGGCTTATTTACTG 2040  
QY 1981 AATTGGACCACTTTTAAATTTTCCCTCTTTTAAACAGTTTACGGCTTATTTACTG 2040  
Db 2041 AATAAATACAAGCAAACTCAAGTTATGTCATACCTTTGGATACGAAGACCATACA 2100  
QY 2041 AATAAATACAAGCAAACTCAAGTTATGTCATACCTTTGGATACGAAGACCATACA 2100  
Db 2101 TAATAACCAACATAACATATACAAAGAAATCTTTTAAACAGTTTACGGCTTATTTAGTC 2160  
QY 2101 TAATAACCAACATAACATATACAAAGAAATCTTTTAAACAGTTTACGGCTTATTTAGTC 2160  
Db 2161 ATTTCAAAAAGTAATCATATATCAAACTAGGCACACACTAAGTTCCCTGATTATTTTGT 2220  
QY 2161 ATTTCAAAAAGTAATCATATATCAAACTAGGCACACACTAAGTTCCCTGATTATTTTGT 2220  
Db 2221 TATAATTTAATAATATCTTTATGAGCCCTATATATTCAAATATATATGTTAAACATGTA 2280  
QY 2221 TATAATTTAATAATATCTTTATGAGCCCTATATATTCAAATATATATGTTAAACATGTA 2280  
Db 2281 TCCATGTTCTTTTCC 2297  
QY 2281 TCCATGTTCTTTTCC 2297

RESULT 3 AR025358 2297 bp DNA PAT 05-DEC-1998  
LOCUS Sequence 1 from patent US 5798239.  
DEFINITION AR025358  
ACCESSION AR025358  
VERSION AR025358.1 GI:3977986  
KEYWORDS  
SOURCE Unknown.  
ORGANISM Unknown.  
REFERENCE Unclassified.  
1 (bases 1 to 2297)  
AUTHORS Wilson,P.J., Morris,C.,Phillip, Anson,D.,Stewart, Occhiodoro,T.,  
Bielicki,J., Clements,P.Roy and Hopwood,J.Joseph.

TITLE Glycosylation variants of iduronate 2-sulfatase  
JOURNAL Patent: US 5798239-A 1 25-AUG-1998;  
FEATURES Location/Qualifiers  
1..2297  
source /organism="unknown"  
BASE COUNT 553 a 600 c 510 g 634 t  
ORIGIN  
Query Match 100.0%; Score 2297; DB 1; Length 2297;  
Best Local Similarity 100.0%; Pred. No. 0.00e+00;  
Matches 2297; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
Db 1 CGGCTGTGTGGCGAGCTTTCATGAGTTCGCGACGAGAGGTCTCTGTGGCTCGCGCGC 60  
QY 1 CGGCTGTGTGGCGAGCTTTCATGAGTTCGCGACGAGAGGTCTCTGTGGCTCGCGCGC 60  
Db 61 TGTCTACTGCGGCACCTGCTGAGCCTGTCCCGCGCGCTCTGAAGCGCGCGCTCGAAGC 120  
QY 61 TGTCTACTGCGGCACCTGCTGAGCCTGTCCCGCGCGCTCTGAAGCGCGCGCTCGAAGC 120  
Db 121 CGAAATGCCGCCACCCGCGGACCGGCGAGGCTTCTCTGGCTGGCTGTGTTCTGAGCTC 180  
QY 121 CGAAATGCCGCCACCCGCGGACCGGCGAGGCTTCTCTGGCTGGCTGTGTTCTGAGCTC 180  
Db 181 CGTCTGCGTGGCCCTCGGATCCGAAACGAGGCGCAACTCGACCACAGATGCTCTGAACGT 240  
QY 181 CGTCTGCGTGGCCCTCGGATCCGAAACGAGGCGCAACTCGACCACAGATGCTCTGAACGT 240  
Db 241 TCTTCTCATCATCGTGGATGACCTGGCGCCCTCCCTGGGCTGTTATGGGATAAAGCTGT 300  
QY 241 TCTTCTCATCATCGTGGATGACCTGGCGCCCTCCCTGGGCTGTTATGGGATAAAGCTGT 300  
Db 301 GAGGTCCCAATATTTGACCACTGTCATCCACAGCTCCTCTTCCAGAAATSCCTTTGC 360  
QY 301 GAGGTCCCAATATTTGACCACTGTCATCCACAGCTCCTCTTCCAGAAATSCCTTTGC 360  
Db 361 GCAGCAAGCAGTGTGCGCCCGCGAGCGCTTTCTTCTTCTCACTGCGAGGAGACTGACAC 420  
QY 361 GCAGCAAGCAGTGTGCGCCCGCGAGCGCTTTCTTCTTCTCACTGCGAGGAGACTGACAC 420  
Db 421 CACCGCGCTGTACGACTTCAACTCTTCTGAGGGTGCACGCTGGAACCTTCTCCACCAT 480  
QY 421 CACCGCGCTGTACGACTTCAACTCTTCTGAGGGTGCACGCTGGAACCTTCTCCACCAT 480  
Db 481 CCCCCAGTACTTCAAGGAGAAATGGCTATGTGACCATGTCTGGGAAAGTCTTTTCCACC 540  
QY 481 CCCCCAGTACTTCAAGGAGAAATGGCTATGTGACCATGTCTGGGAAAGTCTTTTCCACC 540  
Db 541 TGGGATATCTTCTAACCATACCGATGATTCTCGTATAGCTGTTTCCACCTTATCA 600  
QY 541 TGGGATATCTTCTAACCATACCGATGATTCTCGTATAGCTGTTTCCACCTTATCA 600  
Db 601 TCCTTCTCTGAGAAGTATGAAAACACTAAGACATGTCGAGGGCCAGATGGAGAACTCCA 660  
QY 601 TCCTTCTCTGAGAAGTATGAAAACACTAAGACATGTCGAGGGCCAGATGGAGAACTCCA 660  
Db 661 TGCACACCTGCTTTGCCCTGTGGATGTGCTGGATGTTCCCGAGGGCACCTTGCCTGACAA 720  
QY 661 TGCACACCTGCTTTGCCCTGTGGATGTGCTGGATGTTCCCGAGGGCACCTTGCCTGACAA 720  
Db 721 ACAGAGCACTGAGCAAGCCATACAGTTGTTGAAAAGATGAAAACGTCAGCCAGTCTTT 780  
QY 721 ACAGAGCACTGAGCAAGCCATACAGTTGTTGAAAAGATGAAAACGTCAGCCAGTCTTT 780  
Db 781 CTTCTGCGCGTTGGGTATCATAAAGCCACACATCCCTTTCAGATACCCCAAGGAATTTCA 840  
QY 781 CTTCTGCGCGTTGGGTATCATAAAGCCACACATCCCTTTCAGATACCCCAAGGAATTTCA 840  
Db 841 GAAATTGTATCCCTTGGAGAAACATCACCTGCCCCCGGATCCCGAGGTCCCTGATGGCCT 900  
QY 841 GAAATTGTATCCCTTGGAGAAACATCACCTGCCCCCGGATCCCGAGGTCCCTGATGGCCT 900  
Db 901 ACCCCTGTGGCCTACAAACCCCTGGATGACATCAGGCAACGGGAAGACGTCACACCTTT 960

Qy	1981	AATGGGACCAGT	TTTTTTTTTAATTTCCCTCTCTTTTAAACAAGATTACGGCTTATTATTCTG	204
Db	2041	AATAATACAAAGCAAAACAAC	TCAAGTATATGTGCATACCTTTGGATACAGAACCATACA	2100
Qy	2041	AATAATACAAAGCAAAACAAC	TCAAGTATATGTGCATACCTTTGGATACAGAACCATACA	2100
Db	2101	TAATAACCAACATACAT	TATACACAAGAATACCTTCATTATTTGTGGAAATTTAGTGC	2160
Qy	2101	TAATAACCAACATACAT	TATACACAAGAATACCTTCATTATTTGTGGAAATTTAGTGC	2160
Db	2161	ATTTCAAAAAGTAAATCATAT	TATCAAACTAGGCACCACTAAAGTTCCTCGATTATTTTCTT	2220
Qy	2161	ATTTCAAAAAGTAAATCATAT	TATCAAACTAGGCACCACTAAAGTTCCTCGATTATTTTCTT	2220
Db	2221	TATAATTTAATAATATATCT	TATGAGCCCTATATATTTCAAAATATATGTTAAACATGAA	2280
Qy	2221	TATAATTTAATAATATATCT	TATGAGCCCTATATATTTCAAAATATATGTTAAACATGAA	2280
Db	2281	TCCATGTTCTTTTTC	2297	
Qy	2281	TCCATGTTCTTTTTC	2297	
RESULT	4			
LOCUS		HUMIDSALT	1314 bp	mrna
DEFINITION		Homo sapiens iduronate-2-sulphatase (IDS)	mrna, complete cds.	PRI 11-MAR-1996
ACCESSION		L40586		
VERSION		L40586.1	GI:1222629	
KEYWORDS			iduronate-2-sulphatase.	
SOURCE		Homo sapiens		
ORGANISM		Homo sapiens		
REFERENCE			Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.	
AUTHORS			1 (bases 1 to 1314)	
TITLE			Malmgren,H., Carlberg,B.M., Pettersson,U. and Bondeson,M.L. Identification of an alternative transcript from the human iduronate-2-sulfatase (IDS) gene	
JOURNAL			Genomics 29 (1), 291-293 (1995)	
MEDLINE		96079126		
FEATURES			Location/Qualifiers	
source			1..1314	
			/organism="Homo sapiens"	
			/db_xref="taxon:9606"	
			/clone="23.2"	
			/tissue_type="lymphocyte"	
gene			1..1291	
			/gene="IDS"	
exon			1..230	
			/gene="IDS"	
			/note="G00-120-521"	
			/number=1	
CDS			128..1159	
			/gene="IDS"	
			/codon_start=1	
			/db_xref="GDB:G00-120-521"	
			/product="iduronate-2-sulphatase"	
			/protein_id="AA92014.1"	
			/db_xref="GI:1226630"	
			/translation="MPPPTGRGLWLGLVSSVCVALGSGTQANSTTDALNVLDLRLSGYQDKLVRSNPNDQLASHLSLLFQNAQQAVCAPSRVSFLTGRREDFYNSVRYRHAGNSFIPQYFKENGYSVTKVFPHGIGISNHTDDSPSYSMFSFSSBKYEKTKCRPGDELHANLLCPVDLDVPEGLPKQSTEQAIQLLEKMKTFPLAVGKHKPIFPFRYKPEQKLYPLENITLAPDEVPDGLPPVAYNPWMDIRGQALNISVPGPIPVDFQRIKIOSYFASVSYLDTQVGRLLSALDQLQLANSTITAHGFLMRTNT"	
exon			231..367	
			/gene="IDS"	
			/note="G00-120-521"	
			/number=2	
			368..545	
			/gene="IDS"	
exon			/note="G00-120-521"	
			/number=3	

Qy	598	TCATCCTTCCCTGAGAAAGTATGAACACATAAGACATGTGCAGGGGCCAGATGAGAACT	657		
Dd	661	CCATGCCAACCTGCTTTGCCCTGTGGATGTGCTGGATGTTCCTCCGAGGCGACCTTGCCCTGA	720		
Qy	658	CCATGCCAACCTGCTTTGCCCTGTGGATGTGCTGGATGTTCCTCCGAGGCGACCTTGCCCTGA	717		
Dd	721	CAACAGAGCACTGAGCAAGCCATACAGTTGTTGGAAAAAGATGAAAACGTCAGCCAGTCC	780		
Qy	718	CAACAGAGCACTGAGCAAGCCATACAGTTGTTGGAAAAAGATGAAAACGTCAGCCAGTCC	777		
Dd	781	TTTTCTCTCGCCCTGGGTATCATAGCCACACATCCCCTTCAGATTACCACAAGGAATT	840		
Qy	778	TTTTCTCTCGCCCTGGGTATCATAGCCACACATCCCCTTCAGATTACCACAAGGAATT	837		
Dd	841	TCAGAAGTTGATCCCTTGGAGAGACATCACCTCGCCCCCGATCCGAGGTCCCTGATGG	900		
Qy	838	TCAGAAGTTGATCCCTTGGAGAGACATCACCTCGCCCCCGATCCGAGGTCCCTGATGG	897		
Dd	901	CCTACCCCTGTGCCTACAAACCCCTGGATGGACATCAGGCAACGGGAAGACGTCCCAAGC	960		
Qy	898	CCTACCCCTGTGCCTACAAACCCCTGGATGGACATCAGGCAACGGGAAGACGTCCCAAGC	957		
Dd	961	CTTAACATCAGTGTGGGTATGTGCAATTCCTGTGGACTTTTCAGCGGAAAAATCCGCCA	1020		
Qy	958	CTTAACATCAGTGTGGGTATGTGCAATTCCTGTGGACTTTTCAGCGGAAAAATCCGCCA	1017		
Dd	1021	GAGTACTTTGCCCTGTGTCATATTTGGATACACAGCTCGGCCGCTCTTGAGTGCCTTT	1080		
Qy	1018	GAGTACTTTGCCCTGTGTCATATTTGGATACACAGCTCGGCCGCTCTTGAGTGCCTTT	1077		
Dd	1081	GGAGGATCTTCAGCTGGCCAACAGCACCATCATTTACCTCGGATCATGG	1134		
Qy	1078	GGAGGATCTTCAGCTGGCCAACAGCACCATCATTTACCTCGGATCATGG	1131		
RESULT	5	MUSIDS	1750 bp mRNA	ROD	19-AUG-1993
LOCUS		Mus musculus iduronate sulfatase (ids)	mRNA,	complete cds.	
DEFINITION		L07921			
ACCESSION		L07921.1	GI:349713		
VERSION					
KEYWORDS					
SOURCE					
ORGANISM					
REFERENCE					
AUTHORS					
TITLE					
JOURNAL					
MEDLINE					
FEATURES					
source					
gene					
CDS					







```
/tissue_lib="lambda Charon 35, ATCC No. LAOXN101"
/map="Xq27.3-q28"
join(L13321.1:1..536,L13322.1:1..426,L13323.1:1..388,
L13324.1:1..366,L13325.1:1..490,L13326.1:1..379,
L13327.1:1..319,L13328.1:1..404,1..1120)
/gene="IDS"
join(L13321.1:332..434,L13322.1:157..293,L13323.1:82..259,
L13324.1:131..219,L13325.1:125..325,L13326.1:188..258,
L13327.1:99..225,L13328.1:129..302,128..600)
/gene="IDS"
/codon_start=1
/product="iduronate-2-sulfatase"
/protein_id="AAJ16877.1"
/db_xref="GI:405205"
/translacion="MPPRTGRLLWLGVLSSVCVALGSETQANSTFDALNLLIIV
DDLRPSGLKDLVIRPNIDQLASHLLFQNAFAQAVCAPSRVSFLTGRPDTRL
YDFNSYWRVHAGNESTIPOYFKENGIVTMSVGVHPHGIGNSNHTDDSPYFPH
SSEKYENTKCRGPDGELHNLCPVDVLDVECTLPDKOSTEOAIOLLEKMTSASP
FFLAVGYHKPHIPRYPEKFKLYPLENITLAPDPEVDGLPPVAYNPWMIRQREDV
QALNISYPYGPVDFORKIKQSYFASVSYLDTOVQRLLSALDLDLANSTIIAFTSD
HGVALGBHEWAKYSNFDVATHVLPIFYVPGRTASLPEAGEKLFPLDLPFDSASQLE
PGQSMDLVELISLFPFLAGLAGLQVPRCPVPSPHFVELCREGNLKHFRFRLDEE
PVLPGNPRELIAVSOYPRPSDIPQWNSDKPSLKDKIKMGYSIRIDYRYTVWGFNP
EFLANFSDIHAGELYFVDSPLQDHNYNDSQGDLPQLMP"
order(L13328.1:303..404,1..127)
/gene="IDS"
/number=8
128..>1120
/gene="IDS"
/number=9
/evidence-experimental
BASE COUNT 318 a 232 c 212 g 358 t
ORIGIN

Query Match 43.4%; Score 996; DB 6; Length 1120;
Best Local Similarity 100.0%; Pred. No. 0.00e+00;
Matches 996; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Db 125 CAGCAGGCAATCCATGACCTTGTGGAACTTGTCCTCTTTTCCACGCTGGAC 184
QY 1302 CAGCAGGCAATCCATGACCTTGTGGAACTTGTCCTCTTTTCCACGCTGGAC 1361
Db 185 TTGCAGGACTCGAGTTCACCTCGCTGCCCGCTTCTTCATTTCCAGCTGTGCA 244
QY 1362 TTGCAGGACTCGAGTTCACCTCGCTGCCCGCTTCTTCATTTCCAGCTGTGCA 1421
Db 245 GAGAGGCAAGAACTTCTGAGCAATTTTCGATCCGCTGACTTGAAGAGGATCGTACC 304
QY 1422 GAGAGGCAAGAACTTCTGAGCAATTTTCGATCCGCTGACTTGAAGAGGATCGTACC 1481
Db 305 TCCCTGTGTAATCCCGGTGAAGTTCGCTATAGCCAGTATCCCGGCTTCAGACATCC 364
QY 1482 TCCCTGTGTAATCCCGGTGAAGTTCGCTATAGCCAGTATCCCGGCTTCAGACATCC 1541
Db 365 CTCAGTGGAACTTCGACAGCGAGTTTAAAGATATAAGATCATGGGCTATTCATAC 424
QY 1542 CTCAGTGGAACTTCGACAGCGAGTTTAAAGATATAAGATCATGGGCTATTCATAC 1601
Db 425 GCACCATAGACTATAGTATAGTGTGGGTGGTTCATCCCTGATGAATTTCTAGCTA 484
QY 1602 GCACCATAGACTATAGTATAGTGTGGGTGGTTCATCCCTGATGAATTTCTAGCTA 1661
Db 485 ACTTTTCTGACATCCATGACGGGAACGTATTTTGTGGATCTGACCATTCGAGGATC 544
QY 1662 ACTTTTCTGACATCCATGACGGGAACGTATTTTGTGGATCTGACCATTCGAGGATC 1721
Db 545 ACAATATGTAATGATTCACCAAGTGGAGATCTTTTCCAGTGTGATGCTTCAGTTT 604
QY 1722 ACAATATGTAATGATTCACCAAGTGGAGATCTTTTCCAGTGTGATGCTTCAGTTT 1781
Db 605 TGCCAAACCATGGATGGCAAAATGTGATGTGCTCCCTTCCAGCTGGTGAGAGGAGGTAG 664
QY 1782 TGCCAAACCATGGATGGCAAAATGTGATGTGCTCCCTTCCAGCTGGTGAGAGGAGGTAG 1841
```

```
Db 665 AGCTGTCGCTTTTGTGATTACCCATTAATTTGGAAGCAGCCTGAGGCTAGTTAATCCAA 724
QY 1842 AGCTGTCGCTTTTGTGATTACCCATTAATTTGGAAGCAGCCTGAGGCTAGTTAATCCAA 1901
Db 725 ACATGCATCAACAATTTGGCCTCGAGATATGTAACAGCCAAACCTTTTCGTTAGTCTTT 784
QY 1902 ACATGCATCAACAATTTGGCCTCGAGATATGTAACAGCCAAACCTTTTCGTTAGTCTTT 1961
Db 785 ATTAATAATTTATAATTTGTAATTTGGACCACTTTTAAATTTCCCTCTTTTAAAC 844
QY 1962 ATTAATAATTTATAATTTGTAATTTGGACCACTTTTAAATTTCCCTCTTTTAAAC 2021
Db 845 AGTTACGGCTTATTTACTGAATAAATACAAAGCAAACTCAAGTTATGTGCATACCTTT 904
QY 2022 AGTTACGGCTTATTTACTGAATAAATACAAAGCAAACTCAAGTTATGTGCATACCTTT 2081
Db 905 TGGATACGAGAACCATACATAATAACCAACATATATACACAAGAAATACTTTCAT 964
QY 2082 TGGATACGAGAACCATACATAATAACCAACATATATACACAAGAAATACTTTCAT 2141
Db 965 TATTTCTGCAATTTACTGCACTTTCAAAAAGTATATATATCAAACTAGGCACACACTA 1024
QY 2142 TATTTCTGCAATTTACTGCACTTTCAAAAAGTATATATATCAAACTAGGCACACACTA 2201
Db 1025 AGTTCTCTGATTTTGTATTAATTTAATAATATATATCTTATGAGCCCTATATATCAAA 1084
QY 2202 AGTTCTCTGATTTTGTATTAATTTAATAATATATATCTTATGAGCCCTATATATCAAA 2261
Db 1085 ATATATGTTAAACATGTAATCCATGTTCTTTTTC 1120
QY 2262 ATATATGTTAAACATGTAATCCATGTTCTTTTTC 2297
```

```
RESULT 7
LOCUS AR025360 4428 bp DNA PAT 05-DEC-1998
DEFINITION Sequence 6 from patent US 5798239.
ACCESSION AR025360
VERSION AR025360.1 GI:3977988
KEYWORDS
SOURCE Unknown.
ORGANISM Unclassified.
REFERENCE 1 (bases 1 to 4428)
AUTHORS Wilson,P.J., Morris,C.Phillip, Anson,D.Stewart, Occhiodoro,T.,
Bielicki,J., Clements,P.Roy and Hopwood,J.Joseph.
TITLE Glycosylation variants of iduronate 2-sulfatase
JOURNAL Patent: US 5798239-A 6 25-AUG-1998;
FEATURES Location/Qualifiers
source
1..4428
BASE COUNT 1033 a 1064 c 1042 g 1285 t 4 others
ORIGIN

Query Match 43.4%; Score 996; DB 1; Length 4428;
Best Local Similarity 100.0%; Pred. No. 0.00e+00;
Matches 996; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Db 3433 CAGCAGGCAATCCATGAGCCTTGTGAACTTGTGCTCTTTTCCACGCTGGCTGGAC 3492
QY 1302 CAGCAGGCAATCCATGAGCCTTGTGAACTTGTGCTCTTTTCCACGCTGGCTGGAC 1361
Db 3493 TTGCAGGACTCGAGTTCACCTCGCTGCCCGCTTCTTTCATTTACGTTGAGCTGTGCA 3552
QY 1362 TTGCAGGACTCGAGTTCACCTCGCTGCCCGCTTCTTTCATTTACGTTGAGCTGTGCA 1421
Db 3553 GAGAAGGCAAGAACCTTCTGAAGCATTTTCGATTCGCTGACTTGAAGAGGATCCGTACC 3612
QY 1422 GAGAAGGCAAGAACCTTCTGAAGCATTTTCGATTCGCTGACTTGAAGAGGATCCGTACC 1481
Db 3613 TCCCTGGTAAATCCCGTGAACCTGATTCGCTATAGCCAGTATCCCGGCTTCAGACATCC 3672
QY 1482 TCCCTGGTAAATCCCGTGAACCTGATTCGCTATAGCCAGTATCCCGGCTTCAGACATCC 1541
```

Db 3673 CTCAGTGAATTCGACAGCGAGCTTTAAAGATATATAAGATCATGGGCTATTCCATAC 3732  
|||||  
Qy 1542 CTCAGTGAATTCGACAGCGAGCTTTAAAGATATATAAGATCATGGGCTATTCCATAC 1601  
  
Db 3733 GCACCATAGACTATAGGTATAGTGTGGGTGGCTTCAATCCTGATGAATTTCTAGCTA 3792  
|||||  
Qy 1602 GCACCATAGACTATAGGTATAGTGTGGGTGGCTTCAATCCTGATGAATTTCTAGCTA 1661  
  
Db 3793 ACTTTCTGACATCCATGACAGGGGAACCTGATTTTGTGGATCTGACCCATTCGAGATC 3852  
|||||  
Qy 1662 ACTTTCTGACATCCATGACAGGGGAACCTGATTTTGTGGATCTGACCCATTCGAGATC 1721  
  
Db 3853 ACAATATGTAATGATTTCCCAAGGTGGAGATCTTTTCCAGTGTGTGATGCCCTTGAGTTT 3912  
|||||  
Qy 1722 ACAATATGTAATGATTTCCCAAGGTGGAGATCTTTTCCAGTGTGTGATGCCCTTGAGTTT 1781  
  
Db 3913 TGCCAAACCATGGATGGCAAAATGTGATGCTCCCTTCCAGCTGGTGAGAGGAGGAGTTAG 3972  
|||||  
Qy 1782 TGCCAAACCATGGATGGCAAAATGTGATGCTCCCTTCCAGCTGGTGAGAGGAGGAGTTAG 1841  
  
Db 3973 AGCTGGTGGTTTGTGATTTACCCATAATATTGGAAGCAGCCTGAGGGCTAGTTAATCCAA 4032  
|||||  
Qy 1842 AGCTGGTGGTTTGTGATTTACCCATAATATTGGAAGCAGCCTGAGGGCTAGTTAATCCAA 1901  
  
Db 4033 ACATGCATCAACAATTTGGCCTGAGAATATGTAACAGGCCAAACCTTTTCGTTTGTAGTCTT 4092  
|||||  
Qy 1902 ACATGCATCAACAATTTGGCCTGAGAATATGTAACAGGCCAAACCTTTTCGTTTGTAGTCTT 1961  
  
Db 4093 ATTAATAATTTAATTTGGTAAATTTGGACCAAGTGTGTTTAAATTTCCCTCTTTTAAAC 4152  
|||||  
Qy 1962 ATTAATAATTTAATTTGGTAAATTTGGACCAAGTGTGTTTAAATTTCCCTCTTTTAAAC 2021  
  
Db 4153 AGTTACGGCTTATTTACTGAATAAATACAAGCAAAACCTCAAGTTATGTCTATACCTT 4212  
|||||  
Qy 2022 AGTTACGGCTTATTTACTGAATAAATACAAGCAAAACCTCAAGTTATGTCTATACCTT 2081  
  
Db 4213 TGGATACGAAGACCATACATAAATACCAACATAACATATATATATCTTATGAGCCCTATATATTCAAA 4272  
|||||  
Qy 2082 TGGATACGAAGACCATACATAAATACCAACATAACATATATATCTTATGAGCCCTATATATTCAAA 2261  
  
Db 4273 TATTTGTGGAATTTAGTCATTTCAAAAAGTAAATCATATATCAAACTAGGCACCACTA 4332  
|||||  
Qy 2142 TATTTGTGGAATTTAGTCATTTCAAAAAGTAAATCATATATCAAACTAGGCACCACTA 2201  
  
Db 4333 AGTTCCCTGATTTTGTGTTTATATTAATTAATAATATCTTATGAGCCCTATATATTCAAA 4392  
|||||  
Qy 2202 AGTTCCCTGATTTTGTGTTTATATTAATTAATAATATCTTATGAGCCCTATATATTCAAA 2261  
  
Db 4393 ATATATGTTAAACATGTAATCCATGTTTCTTTTCC 4428  
|||||  
Qy 2262 ATATATGTTAAACATGTAATCCATGTTTCTTTTCC 2297

RESULT 8 192329 4428 bp DNA PAT 01-DEC-1998  
LOCUS Sequence 6 from patent US 5728381.  
DEFINITION  
ACCESSION 192329  
VERSION 192329.1 GI:3936799  
KEYWORDS  
SOURCE Unknown.  
ORGANISM Unknown.  
REFERENCE 1 (bases 1 to 4428)  
AUTHORS Wilson, P.J., Morris, C. Phillip, Anson, D. Stewart, Occhiodoro, T., Bielecki, J., Clements, P. Roy and Hopwood, J. Joseph.  
TITLE Glycosylation variants of iduronate 2-sulfatase  
JOURNAL Patent: US 5728381-A 6 17-MAR-1998;  
FEATURES Location/Qualifiers  
source 1..4428  
/organism="unknown"  
BASE COUNT 1033 a 1064 c 1042 g 1285 t 4 others  
ORIGIN

Query Match 43.4%; Score 996; DB 1; Length 4428;  
Best Local Similarity 100.0%; Pred. No. 0.00e+00;  
Matches 996; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
  
Db 3433 CAGCGAGCAATCCATGAGACTTGTGGAACCTGTGTCTCTTTTCCACGCTGGCTGAC 3492  
Qy 1302 CAGCGAGCAATCCATGAGACTTGTGGAACCTGTGTCTCTTTTCCACGCTGGCTGAC 1361  
  
Db 3493 TTGCAGGACTGCGAGTTCCACCTGCGCTGCCCGTTCCTTCATTTTCAGCTGTGAGCTG 3552  
Qy 1362 TTGCAGGACTGCGAGTTCCACCTGCGCTGCCCGTTCCTTCATTTTCAGCTGTGAGCTG 1421  
  
Db 3553 GAGAAGCAAGAACCTTCTGAAGCATTTTGGATTCGCTGATTCGGAAGAGGATCCGTACC 3612  
Qy 1422 GAGAAGCAAGAACCTTCTGAAGCATTTTGGATTCGCTGATTCGGAAGAGGATCCGTACC 1481  
  
Db 3613 TCCTGTGTAATCCCGTGAACCTGATTCGCTATAGCCAGTATCCCGGCTTCAGACATCC 3672  
Qy 1482 TCCTGTGTAATCCCGTGAACCTGATTCGCTATAGCCAGTATCCCGGCTTCAGACATCC 1541  
  
Db 3673 CTCAGTGGAAATCTGCAAGCCGAGTTTAAAGATATAAAGATCATGGGCTATTTCCATAC 3732  
Qy 1542 CTCAGTGGAAATCTGCAAGCCGAGTTTAAAGATATAAAGATCATGGGCTATTTCCATAC 1601  
  
Db 3733 GCACCATAGACTATAGGTATAGTGTGGGTGGCTTCAATCCCTGATGAATTTCTAGCTA 3792  
Qy 1602 GCACCATAGACTATAGGTATAGTGTGGGTGGCTTCAATCCCTGATGAATTTCTAGCTA 1661  
  
Db 3793 ACTTTTCTGACATCCATGAGGGGAACCTGATTTTGTGGATTCGACCCATTCGAGGATC 3852  
Qy 1662 ACTTTTCTGACATCCATGAGGGGAACCTGATTTTGTGGATTCGACCCATTCGAGGATC 1721  
  
Db 3853 ACAATATGTAATGATTTCCCAAGTGGAGATCTTTTCCAGTGTGTGATGCCCTTGAGTTT 3912  
Qy 1722 ACAATATGTAATGATTTCCCAAGTGGAGATCTTTTCCAGTGTGTGATGCCCTTGAGTTT 1781  
  
Db 3913 TGGCAACCATGGATGGCAAAATGTGATGCTCCCTTCCAGCTGGTGAGAGGAGGAGTTAG 3972  
Qy 1782 TGGCAACCATGGATGGCAAAATGTGATGCTCCCTTCCAGCTGGTGAGAGGAGGAGTTAG 1841  
  
Db 3973 AGCTGGTGGTTTGTGATTTACCCATAATATTGGAAGCAGCCTGAGGGCTAGTTAATCCAA 4032  
Qy 1842 AGCTGGTGGTTTGTGATTTACCCATAATATTGGAAGCAGCCTGAGGGCTAGTTAATCCAA 1901  
  
Db 4033 ACATGCATCAACAATTTGGCTGAGAATATGTAACAGCCAAACCTTTTTCGTTTGTAGTCTT 4092  
Qy 1902 ACATGCATCAACAATTTGGCTGAGAATATGTAACAGCCAAACCTTTTTCGTTTGTAGTCTT 1961  
  
Db 4093 ATTAATAATTTAATTTGGTAAATTTGGACCAAGTGTGTTTAAATTTCCCTCTTTTAAAC 4152  
Qy 1962 ATTAATAATTTAATTTGGTAAATTTGGACCAAGTGTGTTTAAATTTCCCTCTTTTAAAC 2021  
  
Db 4153 AGTTACGGCTTATTTACTGAATAAATACAAGCAAAACCTCAAGTTATGTCTATACCTT 4212  
Qy 2022 AGTTACGGCTTATTTACTGAATAAATACAAGCAAAACCTCAAGTTATGTCTATACCTT 2081  
  
Db 4213 TGGATACGAAGACCATACATAAATACCAACATAACATATATATCTTATGAGCCCTATATATTCAAA 4272  
Qy 2082 TGGATACGAAGACCATACATAAATACCAACATAACATATATATCTTATGAGCCCTATATATTCAAA 2141  
  
Db 4273 TATTTGTGGAATTTAGTCATTTCAAAAAGTAAATCATATATCAAACTAGGCACCACTA 4332  
Qy 2142 TATTTGTGGAATTTAGTCATTTCAAAAAGTAAATCATATATCAAACTAGGCACCACTA 2201  
  
Db 4333 AGTTCCCTGATTTTGTGTTTATATTAATTAATAATATCTTATGAGCCCTATATATTCAAA 4392  
Qy 2202 AGTTCCCTGATTTTGTGTTTATATTAATTAATAATATCTTATGAGCCCTATATATTCAAA 2261  
  
Db 4393 ATATATGTTAAACATGTAATCCATGTTTCTTTTCC 4428  
Qy 2262 ATATATGTTAAACATGTAATCCATGTTTCTTTTCC 2297

RESULT 9 HUMIDS 36845 bp DNA PRI 16-AUG-1994  
LOCUS Homo sapiens iduronate sulphate sulphatase (IDS) gene, complete  
DEFINITION cds.  
ACCESSION L35485  
VERSION L35485.1 GI:530140  
KEYWORDS iduronate sulphate sulphatase.  
SOURCE Homo sapiens DNA.  
ORGANISM Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;  
Eutheria; Primates; Catarrhini; Hominidae; Homo.  
REFERENCE 1 (bases 1 to 36845)  
AUTHORS Lu,F., Lu,J., Clingman,R.L., Wentland,M.A., Muzny,D.M., Gu,Y.,  
Nelson,D.L. and Gibbs,R.A.  
TITLE Complete DNA sequence of the human iduronate sulphate sulphatase  
(ids) locus  
JOURNAL Unpublished (1994)  
FEATURES  
source location/Qualifiers  
1..36845  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/map="Xq27.3-q28"  
join(1057..1159,1901..2037,2704..2881,5155..5243,  
7885..8085,9676..9846,15752..15878,19094..19267,  
22974..23446)  
/gene="IDS"  
exon <1057..1159  
/gene="IDS"  
number=1  
join(1057..1159,1901..2037,2704..2881,5155..5243,  
7885..8085,9676..9846,15752..15878,19094..19267,  
22974..23446)  
/gene="IDS"  
CDS /codon\_start=1  
/product="iduronate sulphate sulphatase"  
/protein\_id="AAA36037.1"  
/db\_xref="GI:530141"  
/translation="MPPPTGRGLLGLVSSVCVALGSETQANSTTDAINLVLLIIV  
DRLRSLGCGDKLVRSPNIDOLASHSLFQNAFAQAQVAPSRVSFLTGRPDTRL  
YDFNSYRWVHAGNESTIPQYFKENGYVTMSGVKVFHGIGSNHDDSDSYWSFPPYHP  
SSEYENKTKCRGDELHANLLCPVDLDVPEGLPKDQTEQAQLLEKMKTSASP  
FFLAVGHYPKIPRYKFKQKLYPLENITLAPDPEVDPGLPPVAYNPWMDIQRDVE  
QALNISVPYGPDPVDFQKIQRSYFASVYLDTOVGRLLSALDLDLANSTIIAFTSD  
HGWALGHEWAKVNSVDVATHVPLIFVVPGRASLPEAGEKLPYLPDPDSASOLME  
PGRQMDLVLESPFLAGLAGLOVPPRCPVPSPFHVLCREGKRLKHFRFRDLEED  
PYLPGNRELAYSQYPRPSDIPQWNSDKPSLKDIIKIMYSIRTIDYRTYVWVGPNPD  
EFLANFSDIHAGELYFVDSPLQDHNMYNDSQGGDLFQLLMP"  
/gene="IDS"  
intron 1160..1900  
/number=1  
exon 1901..2037  
/gene="IDS"  
intron 2038..2703  
/gene="IDS"  
exon 2704..2881  
/gene="IDS"  
intron 2882..5154  
/gene="IDS"  
exon 5155..5243  
/gene="IDS"  
intron 5244..7884  
/gene="IDS"  
exon 7885..8085  
/number=4  
repeat\_region 8086..9675  
/rpt\_family="Alu"  
repeat\_region 9676..9846  
/rpt\_family="L1nel1"

7885..8085  
/gene="IDS"  
number=5  
8086..9675  
/gene="IDS"  
number=5  
9676..9846  
/gene="IDS"  
number=6  
9847..15751  
/gene="IDS"  
number=6  
15752..15878  
/gene="IDS"  
number=7  
15879..19093  
/gene="IDS"  
number=7  
19094..19267  
/gene="IDS"  
number=8  
19268..22972  
/gene="IDS"  
number=8  
22973..22974  
/rpt\_family="Alu"  
22975..22976  
/rpt\_family="L1nel1"  
22977..22978  
/rpt\_family="L1nel1"  
22979..22980  
/rpt\_family="L1nel1"  
22981..22982  
/gene="IDS"  
number=9  
22983..22984  
/rpt\_family="Alu"  
22985..22986  
/rpt\_family="Alu"  
22987..22988  
/rpt\_family="Alu"  
22989..22990  
/rpt\_family="Alu"  
22991..22992  
/rpt\_family="Alu"  
22993..22994  
/rpt\_family="Alu"  
22995..22996  
/rpt\_family="Alu"  
22997..22998  
/rpt\_family="Alu"  
22999..23000  
/rpt\_family="Alu"  
23001..23002  
/rpt\_family="Alu"  
23003..23004  
/rpt\_family="Alu"  
23005..23006  
/rpt\_family="Alu"  
23007..23008  
/rpt\_family="Alu"  
23009..23010  
/rpt\_family="Alu"  
23011..23012  
/rpt\_family="Alu"  
23013..23014  
/rpt\_family="Alu"  
23015..23016  
/rpt\_family="Alu"  
23017..23018  
/rpt\_family="Alu"  
23019..23020  
/rpt\_family="Alu"  
23021..23022  
/rpt\_family="Alu"  
23023..23024  
/rpt\_family="Alu"  
23025..23026  
/rpt\_family="Alu"  
23027..23028  
/rpt\_family="Alu"  
23029..23030  
/rpt\_family="Alu"  
23031..23032  
/rpt\_family="Alu"  
23033..23034  
/rpt\_family="Alu"  
23035..23036  
/rpt\_family="Alu"  
23037..23038  
/rpt\_family="Alu"  
23039..23040  
/rpt\_family="Alu"  
23041..23042  
/rpt\_family="Alu"  
23043..23044  
/rpt\_family="Alu"  
23045..23046  
/rpt\_family="Alu"  
23047..23048  
/rpt\_family="Alu"  
23049..23050  
/rpt\_family="Alu"  
23051..23052  
/rpt\_family="Alu"  
23053..23054  
/rpt\_family="Alu"  
23055..23056  
/rpt\_family="Alu"  
23057..23058  
/rpt\_family="Alu"  
23059..23060  
/rpt\_family="Alu"  
23061..23062  
/rpt\_family="Alu"  
23063..23064  
/rpt\_family="Alu"  
23065..23066  
/rpt\_family="Alu"  
23067..23068  
/rpt\_family="Alu"  
23069..23070  
/rpt\_family="Alu"  
23071..23072  
/rpt\_family="Alu"  
23073..23074  
/rpt\_family="Alu"  
23075..23076  
/rpt\_family="Alu"  
23077..23078  
/rpt\_family="Alu"  
23079..23080  
/rpt\_family="Alu"  
23081..23082  
/rpt\_family="Alu"  
23083..23084  
/rpt\_family="Alu"  
23085..23086  
/rpt\_family="Alu"  
23087..23088  
/rpt\_family="Alu"  
23089..23090  
/rpt\_family="Alu"  
23091..23092  
/rpt\_family="Alu"  
23093..23094  
/rpt\_family="Alu"  
23095..23096  
/rpt\_family="Alu"  
23097..23098  
/rpt\_family="Alu"  
23099..23100  
/rpt\_family="Alu"  
23101..23102  
/rpt\_family="Alu"  
23103..23104  
/rpt\_family="Alu"  
23105..23106  
/rpt\_family="Alu"  
23107..23108  
/rpt\_family="Alu"  
23109..23110  
/rpt\_family="Alu"  
23111..23112  
/rpt\_family="Alu"  
23113..23114  
/rpt\_family="Alu"  
23115..23116  
/rpt\_family="Alu"  
23117..23118  
/rpt\_family="Alu"  
23119..23120  
/rpt\_family="Alu"  
23121..23122  
/rpt\_family="Alu"  
23123..23124  
/rpt\_family="Alu"  
23125..23126  
/rpt\_family="Alu"  
23127..23128  
/rpt\_family="Alu"  
23129..23130  
/rpt\_family="Alu"  
23131..23132  
/rpt\_family="Alu"  
23133..23134  
/rpt\_family="Alu"  
23135..23136  
/rpt\_family="Alu"  
23137..23138  
/rpt\_family="Alu"  
23139..23140  
/rpt\_family="Alu"  
23141..23142  
/rpt\_family="Alu"  
23143..23144  
/rpt\_family="Alu"  
23145..23146  
/rpt\_family="Alu"  
23147..23148  
/rpt\_family="Alu"  
23149..23150  
/rpt\_family="Alu"  
23151..23152  
/rpt\_family="Alu"  
23153..23154  
/rpt\_family="Alu"  
23155..23156  
/rpt\_family="Alu"  
23157..23158  
/rpt\_family="Alu"  
23159..23160  
/rpt\_family="Alu"  
23161..23162  
/rpt\_family="Alu"  
23163..23164  
/rpt\_family="Alu"  
23165..23166  
/rpt\_family="Alu"  
23167..23168  
/rpt\_family="Alu"  
23169..23170  
/rpt\_family="Alu"  
23171..23172  
/rpt\_family="Alu"  
23173..23174  
/rpt\_family="Alu"  
23175..23176  
/rpt\_family="Alu"  
23177..23178  
/rpt\_family="Alu"  
23179..23180  
/rpt\_family="Alu"  
23181..23182  
/rpt\_family="Alu"  
23183..23184  
/rpt\_family="Alu"  
23185..23186  
/rpt\_family="Alu"  
23187..23188  
/rpt\_family="Alu"  
23189..23190  
/rpt\_family="Alu"  
23191..23192  
/rpt\_family="Alu"  
23193..23194  
/rpt\_family="Alu"  
23195..23196  
/rpt\_family="Alu"  
23197..23198  
/rpt\_family="Alu"  
23199..23200  
/rpt\_family="Alu"  
23201..23202  
/rpt\_family="Alu"  
23203..23204  
/rpt\_family="Alu"  
23205..23206  
/rpt\_family="Alu"  
23207..23208  
/rpt\_family="Alu"  
23209..23210  
/rpt\_family="Alu"  
23211..23212  
/rpt\_family="Alu"  
23213..23214  
/rpt\_family="Alu"  
23215..23216  
/rpt\_family="Alu"  
23217..23218  
/rpt\_family="Alu"  
23219..23220  
/rpt\_family="Alu"  
23221..23222  
/rpt\_family="Alu"  
23223..23224  
/rpt\_family="Alu"  
23225..23226  
/rpt\_family="Alu"  
23227..23228  
/rpt\_family="Alu"  
23229..23230  
/rpt\_family="Alu"  
23231..23232  
/rpt\_family="Alu"  
23233..23234  
/rpt\_family="Alu"  
23235..23236  
/rpt\_family="Alu"  
23237..23238  
/rpt\_family="Alu"  
23239..23240  
/rpt\_family="Alu"  
23241..23242  
/rpt\_family="Alu"  
23243..23244  
/rpt\_family="Alu"  
23245..23246  
/rpt\_family="Alu"  
23247..23248  
/rpt\_family="Alu"  
23249..23250  
/rpt\_family="Alu"  
23251..23252  
/rpt\_family="Alu"  
23253..23254  
/rpt\_family="Alu"  
23255..23256  
/rpt\_family="Alu"  
23257..23258  
/rpt\_family="Alu"  
23259..23260  
/rpt\_family="Alu"  
23261..23262  
/rpt\_family="Alu"  
23263..23264  
/rpt\_family="Alu"  
23265..23266  
/rpt\_family="Alu"  
23267..23268  
/rpt\_family="Alu"  
23269..23270  
/rpt\_family="Alu"  
23271..23272  
/rpt\_family="Alu"  
23273..23274  
/rpt\_family="Alu"  
23275..23276  
/rpt\_family="Alu"  
23277..23278  
/rpt\_family="Alu"  
23279..23280  
/rpt\_family="Alu"  
23281..23282  
/rpt\_family="Alu"  
23283..23284  
/rpt\_family="Alu"  
23285..23286  
/rpt\_family="Alu"  
23287..23288  
/rpt\_family="Alu"  
23289..23290  
/rpt\_family="Alu"  
23291..23292  
/rpt\_family="Alu"  
23293..23294  
/rpt\_family="Alu"  
23295..23296  
/rpt\_family="Alu"  
23297..23298  
/rpt\_family="Alu"  
23299..23300  
/rpt\_family="Alu"  
23301..23302  
/rpt\_family="Alu"  
23303..23304  
/rpt\_family="Alu"  
23305..23306  
/rpt\_family="Alu"  
23307..23308  
/rpt\_family="Alu"  
23309..23310  
/rpt\_family="Alu"  
23311..23312  
/rpt\_family="Alu"  
23313..23314  
/rpt\_family="Alu"  
23315..23316  
/rpt\_family="Alu"  
23317..23318  
/rpt\_family="Alu"  
23319..23320  
/rpt\_family="Alu"  
23321..23322  
/rpt\_family="Alu"  
23323..23324  
/rpt\_family="Alu"  
23325..23326  
/rpt\_family="Alu"  
23327..23328  
/rpt\_family="Alu"  
23329..23330  
/rpt\_family="Alu"  
23331..23332  
/rpt\_family="Alu"  
23333..23334  
/rpt\_family="Alu"  
23335..23336  
/rpt\_family="Alu"  
23337..23338  
/rpt\_family="Alu"  
23339..23340  
/rpt\_family="Alu"  
23341..23342  
/rpt\_family="Alu"  
23343..23344  
/rpt\_family="Alu"  
23345..23346  
/rpt\_family="Alu"  
23347..23348  
/rpt\_family="Alu"  
23349..23350  
/rpt\_family="Alu"  
23351..23352  
/rpt\_family="Alu"  
23353..23354  
/rpt\_family="Alu"  
23355..23356  
/rpt\_family="Alu"  
23357..23358  
/rpt\_family="Alu"  
23359..23360  
/rpt\_family="Alu"  
23361..23362  
/rpt\_family="Alu"  
23363..23364  
/rpt\_family="Alu"  
23365..23366  
/rpt\_family="Alu"  
23367..23368  
/rpt\_family="Alu"  
23369..23370  
/rpt\_family="Alu"  
23371..23372  
/rpt\_family="Alu"  
23373..23374  
/rpt\_family="Alu"  
23375..23376  
/rpt\_family="Alu"  
23377..23378  
/rpt\_family="Alu"  
23379..23380  
/rpt\_family="Alu"  
23381..23382  
/rpt\_family="Alu"  
23383..23384  
/rpt\_family="Alu"  
23385..23386  
/rpt\_family="Alu"  
23387..23388  
/rpt\_family="Alu"  
23389..23390  
/rpt\_family="Alu"  
23391..23392  
/rpt\_family="Alu"  
23393..23394  
/rpt\_family="Alu"  
23395..23396  
/rpt\_family="Alu"  
23397..23398  
/rpt\_family="Alu"  
23399..23400  
/rpt\_family="Alu"  
23401..23402  
/rpt\_family="Alu"  
23403..23404  
/rpt\_family="Alu"  
23405..23406  
/rpt\_family="Alu"  
23407..23408  
/rpt\_family="Alu"  
23409..23410  
/rpt\_family="Alu"  
23411..23412  
/rpt\_family="Alu"  
23413..23414  
/rpt\_family="Alu"  
23415..23416  
/rpt\_family="Alu"  
23417..23418  
/rpt\_family="Alu"  
23419..23420  
/rpt\_family="Alu"  
23421..23422  
/rpt\_family="Alu"  
23423..23424  
/rpt\_family="Alu"  
23425..23426  
/rpt\_family="Alu"  
23427..23428  
/rpt\_family="Alu"  
23429..23430  
/rpt\_family="Alu"  
23431..23432  
/rpt\_family="Alu"  
23433..23434  
/rpt\_family="Alu"  
23435..23436  
/rpt\_family="Alu"  
23437..23438  
/rpt\_family="Alu"  
23439..23440  
/rpt\_family="Alu"  
23441..23442  
/rpt\_family="Alu"  
23443..23444  
/rpt\_family="Alu"  
23445..23446  
/rpt\_family="Alu"  
23447..23448  
/rpt\_family="Alu"  
23449..23450  
/rpt\_family="Alu"  
23451..23452  
/rpt\_family="Alu"  
23453..23454  
/rpt\_family="Alu"  
23455..23456  
/rpt\_family="Alu"  
23457..23458  
/rpt\_family="Alu"  
23459..23460  
/rpt\_family="Alu"  
23461..23462  
/rpt\_family="Alu"  
23463..23464  
/rpt\_family="Alu"  
23465..23466  
/rpt\_family="Alu"  
23467..23468  
/rpt\_family="Alu"  
23469..23470  
/rpt\_family="Alu"  
23471..23472  
/rpt\_family="Alu"  
23473..23474  
/rpt\_family="Alu"  
23475..23476  
/rpt\_family="Alu"  
23477..23478  
/rpt\_family="Alu"  
23479..23480  
/rpt\_family="Alu"  
23481..23482  
/rpt\_family="Alu"  
23483..23484  
/rpt\_family="Alu"  
23485..23486  
/rpt\_family="Alu"  
23487..23488  
/rpt\_family="Alu"  
23489..23490  
/rpt\_family="Alu"  
23491..23492  
/rpt\_family="Alu"  
23493..23494  
/rpt\_family="Alu"  
23495..23496  
/rpt\_family="Alu"  
23497..23498  
/rpt\_family="Alu"  
23499..23500  
/rpt\_family="Alu"  
23501..23502  
/rpt\_family="Alu"  
23503..23504  
/rpt\_family="Alu"  
23505..23506  
/rpt\_family="Alu"  
23507..23508  
/rpt\_family="Alu"  
23509..23510  
/rpt\_family="Alu"  
23511..23512  
/rpt\_family="Alu"  
23513..23514  
/rpt\_family="Alu"  
23515..23516  
/rpt\_family="Alu"  
23517..23518  
/rpt\_family="Alu"  
23519..23520  
/rpt\_family="Alu"  
23521..23522  
/rpt\_family="Alu"  
23523..23524  
/rpt\_family="Alu"  
23525..23526  
/rpt\_family="Alu"  
23527..23528  
/rpt\_family="Alu"  
23529..23530  
/rpt\_family="Alu"  
23531..23532  
/rpt\_family="Alu"  
23533..23534  
/rpt\_family="Alu"  
23535..23536  
/rpt\_family="Alu"  
23537..23538  
/rpt\_family="Alu"  
23539..23540  
/rpt\_family="Alu"  
23541..23542  
/rpt\_family="Alu"  
23543..23544  
/rpt\_family="Alu"  
23545..23546  
/rpt\_family="Alu"  
23547..23548  
/rpt\_family="Alu"  
23549..23550  
/rpt\_family="Alu"  
23551..23552  
/rpt\_family="Alu"  
23553..23554  
/rpt\_family="Alu"  
23555..23556  
/rpt\_family="Alu"  
23557..23558  
/rpt\_family="Alu"  
23559..23560  
/rpt\_family="Alu"  
23561..23562  
/rpt\_family="Alu"  
23563..23564  
/rpt\_family="Alu"  
23565..23566  
/rpt\_family="Alu"  
23567..23568  
/rpt\_family="Alu"  
23569..23570  
/rpt\_family="Alu"  
23571..23572  
/rpt\_family="Alu"  
23573..23574  
/rpt\_family="Alu"  
23575..23576  
/rpt\_family="Alu"  
23577..23578  
/rpt\_family="Alu"  
23579..23580  
/rpt\_family="Alu"  
23581..23582  
/rpt\_family="Alu"  
23583..23584  
/rpt\_family="Alu"  
23585..23586  
/rpt\_family="Alu"  
23587..23588  
/rpt\_family="Alu"  
23589..23590  
/rpt\_family="Alu"  
23591..23592  
/rpt\_family="Alu"  
23593..23594  
/rpt\_family="Alu"  
23595..23596  
/rpt\_family="Alu"  
23597..23598  
/rpt\_family="Alu"  
23599..23600  
/rpt\_family="Alu"  
23601..23602  
/rpt\_family="Alu"  
23603..23604  
/rpt\_family="Alu"  
23605..23606  
/rpt\_family="Alu"  
23607..23608  
/rpt\_family="Alu"  
23609..23610  
/rpt\_family="Alu"  
23611..23612  
/rpt\_family="Alu"  
23613..23614  
/rpt\_family="Alu"  
23615..23616  
/rpt\_family="Alu"  
23617..23618  
/rpt\_family="Alu"  
23619..23620  
/rpt\_family="Alu"  
23621..23622  
/rpt\_family="Alu"  
23623..23624  
/rpt\_family="Alu"  
23625..23626  
/rpt\_family="Alu"  
23627..23628  
/rpt\_family="Alu"  
23629..23630  
/rpt\_family="Alu"  
23631..23632  
/rpt\_family="Alu"  
23633..23634  
/rpt\_family="Alu"  
23635..23636  
/rpt\_family="Alu"  
23637..23638  
/rpt\_family="Alu"  
23639..23640  
/rpt\_family="Alu"  
23641..23642  
/rpt\_family="Alu"  
23643..23644  
/rpt\_family="Alu"  
23645..23646  
/rpt\_family="Alu"  
23647..23648  
/rpt\_family="Alu"  
23649..23650  
/rpt\_family="Alu"  
23651..23652  
/rpt\_family="Alu"  
23653..23654  
/rpt\_family="Alu"  
23655..23656  
/rpt\_family="Alu"  
23657..23658  
/rpt\_family="Alu"  
23659..23660  
/rpt\_family="Alu"  
23661..23662  
/rpt\_family="Alu"  
23663..23664  
/rpt\_family="Alu"  
23665..23666  
/rpt\_family="Alu"  
23667..23668  
/rpt\_family="Alu"  
23669..23670  
/rpt\_family="Alu"  
23671..23672  
/rpt\_family="Alu"  
23673..23674  
/rpt\_family="Alu"  
23675..23676  
/rpt\_family="Alu"  
23677..23678  
/rpt\_family="Alu"  
23679..23680  
/rpt\_family="Alu"  
23681..23682  
/rpt\_family="Alu"  
23683..23684  
/rpt\_family="Alu"  
23685..23686  
/rpt\_family="Alu"  
23687..23688  
/rpt\_family="Alu"  
23689..23690  
/rpt\_family="Alu"  
23691..23692  
/rpt\_family="Alu"  
23693..23694  
/rpt\_family="Alu"  
23695..23696  
/rpt\_family="Alu"  
23697..23698  
/rpt\_family="Alu"  
23699..23700  
/rpt\_family="Alu"  
23701..23702  
/rpt\_family="Alu"  
23703..23704  
/rpt\_family="Alu"  
23705..23706  
/rpt\_family="Alu"  
23707..23708  
/rpt\_family="Alu"  
23709..23710  
/rpt\_family="Alu"  
23711..23712  
/rpt\_family="Alu"  
23713..23714  
/rpt\_family="Alu"  
23715..23716  
/rpt\_family="Alu"  
23717..23718  
/rpt\_family="Alu"  
23719..23720  
/rpt\_family="Alu"  
23721..23722  
/rpt\_family="Alu"  
23723..23724  
/rpt\_family="Alu"  
23725..23726  
/rpt\_family="Alu"  
23727..23728  
/rpt\_family="Alu"  
23729..23730  
/rpt\_family="Alu"  
23731..23732  
/rpt\_family="Alu"  
23733..23734  
/rpt\_family="Alu"  
23735..23736  
/rpt\_family="Alu"  
23737..23738  
/rpt\_family="Alu"  
23739..23740  
/rpt\_family="Alu"  
23741..23742  
/rpt\_family="Alu"  
23743..23744  
/rpt\_family="Alu"  
23745..23746  
/rpt\_family="Alu"  
23747..23748  
/rpt\_family="Alu"  
23749..23750  
/rpt\_family="Alu"  
23751..23752  
/rpt\_family="Alu"  
23753..23754  
/rpt\_family="Alu"  
23755..23756  
/rpt\_family="Alu"  
23757..23758  
/rpt\_family="Alu"  
23759..23760  
/rpt\_family="Alu"  
23761..23762  
/rpt\_family="Alu"  
23763..23764  
/rpt\_family="Alu"  
23765..23766  
/rpt\_family="Alu"  
23767..23768  
/rpt\_family="Alu"  
23769..23770  
/rpt\_family="Alu"  
23771..23772  
/rpt\_family="Alu"  
23773..23774  
/rpt\_family="Alu"  
23775..23776  
/rpt\_family="Alu"  
23777..23778  
/rpt\_family="Alu"  
23779..23780  
/rpt\_family="Alu"  
23781..23782  
/rpt\_family="Alu"  
23783..23784  
/rpt\_family="Alu"  
23785..23786  
/rpt\_family="Alu"  
23787..23788  
/rpt\_family="Alu"  
23789..23790  
/rpt\_family="Alu"  
23791..23792  
/rpt\_family="Alu"  
23793..23794  
/rpt\_family="Alu"  
23795..23796  
/rpt\_family="Alu"  
23797..23798  
/rpt\_family="Alu"  
23799..23800  
/rpt\_family="Alu"  
23801..23802  
/rpt\_family="Alu"  
23803..23804  
/rpt\_family="Alu"  
23805..23806  
/rpt\_family="Alu"  
23807..23808  
/rpt\_family="Alu"  
23809..23810  
/rpt\_family="Alu"  
23811..23812  
/rpt\_family="Alu"  
23813..23814  
/rpt\_family="Alu"  
23815..23816  
/rpt\_family="Alu"  
23817..23818  
/rpt\_family="Alu"  
23819..23820  
/rpt\_family="Alu"  
23821..23822  
/rpt\_family="Alu"  
23823..23824  
/rpt\_family="Alu"  
23825..23826  
/rpt\_family="Alu"  
23827..23828  
/rpt\_family="Alu"  
23829..23830  
/rpt\_family="Alu"  
23831..23832  
/rpt\_family="Alu"  
23833..23834  
/rpt\_family="Alu"  
23835..23836  
/rpt\_family="Alu"  
23837..23838  
/rpt\_family="Alu"  
23839..23840  
/rpt\_family="Alu"  
23841..23842  
/rpt\_family="Alu"  
23843..23844  
/rpt\_family="Alu"  
23845..23846  
/rpt\_family="Alu"  
23847..23848  
/rpt\_family="Alu"  
23849..23850  
/rpt\_family="Alu"  
23851..23852  
/rpt\_family="Alu"  
23853..23854  
/rpt\_family="Alu"  
23855..23856  
/rpt\_family="Alu"  
23857..23858  
/rpt\_family="Alu"  
23859..23860  
/rpt\_family="Alu"  
23861..23862  
/rpt\_family="Alu"  
23863..23864  
/rpt\_family="Alu"  
23865..23866  
/rpt\_family="Alu"  
23867..23868  
/rpt\_family="Alu"  
23869..23870  
/rpt\_family="Alu"  
23871..23872  
/rpt\_family="Alu"  
23873..23874  
/rpt\_family="Alu"  
23875..23876  
/rpt\_family="Alu"  
23877..23878  
/rpt\_family="Alu"  
23879..23880  
/rpt\_family="Alu"  
23881..23882  
/rpt\_family="Alu"  
23883..23884  
/rpt\_family="Alu"  
23885..23886  
/rpt\_family="Alu"  
23887..23888  
/rpt\_family="Alu"  
23889..23890  
/rpt\_family="Alu"  
23891..23892  
/rpt\_family="Alu"  
23893..23894  
/rpt\_family="Alu"  
23895..23896  
/rpt\_family="Alu"  
23897..23898  
/rpt\_family="Alu"  
23899..23900  
/rpt\_family="Alu"  
23901..23902  
/rpt\_family="Alu"  
23903..23904  
/rpt\_family="Alu"  
23905..23906  
/rpt\_family="Alu"  
23907..23908  
/rpt\_family="Alu"  
23909..23910  
/rpt\_family="Alu"  
23911..23912  
/rpt\_family="Alu"  
23913..23914  
/r

Db	23331	ACTTTCTGACATCCATGCGAGGGAACCTGTAATTTGTGGATCTCGACCATTCGAGGATC	23390
Qy	1662	ACTTTCTGACATCCATGCGAGGGAACCTGTAATTTGTGGATCTCGACCATTCGAGGATC	1721
Db	23391	ACAATATGTAATGATCCCAAGGTGGAGATCTTTCCAGTCTTGATGCCCTTGAGTTT	23450
Qy	1722	ACAATATGTAATGATCCCAAGGTGGAGATCTTTCCAGTCTTGATGCCCTTGAGTTT	1781
Db	23451	TGCCAACCATCGATGGCAAAATGTGATGTGCCCTTCACGCTGCTGAGAGGAGGATAG	23510
Qy	1782	TGCCAACCATCGATGGCAAAATGTGATGTGCCCTTCACGCTGCTGAGAGGAGGATAG	1841
Db	23511	AGCTGGTGGTTTTGTTGATATACCCATAATATTGGAAGCAGCCTGAGGGCTAGTTAATCCAA	23570
Qy	1842	AGCTGGTGGTTTTGTTGATATACCCATAATATTGGAAGCAGCCTGAGGGCTAGTTAATCCAA	1901
Db	23571	ACATGCATCAACAATTTGGCGCTGAGATATGTACAGCAGCAACCTTTTCGTTTAGTCTTT	23630
Qy	1902	ACATGCATCAACAATTTGGCGCTGAGATATGTACAGCAGCAACCTTTTCGTTTAGTCTTT	1961
Db	23631	ATTTAAATTTATAATTTGTGTAATTCGACACAGTTTTTTTTTAAATTTCCCTCTTTTAAAC	23690
Qy	1962	ATTTAAATTTATAATTTGTGTAATTCGACACAGTTTTTTTTTAAATTTCCCTCTTTTAAAC	2021
Db	23691	AGTTACGGCTTATTACTGTAATAATACAAAGCAAAACAACTCAAGTTATGTCATACCTT	23750
Qy	2022	AGTTACGGCTTATTACTGTAATAATACAAAGCAAAACAACTCAAGTTATGTCATACCTT	2081
Db	23751	TGGATACGAAGACCATACATTAATACCAACATTAACATTAATACCAAGAAATCTTCAT	23810
Qy	2082	TGGATACGAAGACCATACATTAATACCAACATTAACATTAATACCAAGAAATCTTCAT	2141
Db	23811	TATTGTGGAAATTTAGTGCATTTCAAAAAGTAATCATATATCAAACTAGGCACACACTA	23870
Qy	2142	TATTGTGGAAATTTAGTGCATTTCAAAAAGTAATCATATATCAAACTAGGCACACACTA	2201
Db	23871	AGTTCCCTGATTATTTGTTTAAATTAATATATCTTTAGGCCCTATATATTCAAA	23930
Qy	2202	AGTTCCCTGATTATTTGTTTAAATTAATATATCTTTAGGCCCTATATATTCAAA	2261
Db	23931	ATATTATGTTAAACATGTAATCCATGTTTCTTTTTC	23965
Qy	2262	ATATTATGTTAAACATGTAATCCATGTTTCTTTTTC	2296
RESULT	10		
LOCUS	AF011889	326663 bp	DNA
DEFINITION	Human Xq28 cosmids U136G1, U142F2, U6986, U145C10, U169A5, U84H1, U24D12, U80A7, U153E6, L35485, and R7-163A8 containing iduronate 2-sulfatase gene and pseudogene, complete sequence.		
ACCESSION	AF011889	U78816	U66082
VERSION	AF011889.1	GI:2335186	
KEYWORDS	HTG.		
SOURCE	human.		
ORGANISM	Homo sapiens		
REFERENCE	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.		
AUTHORS	1 (bases 195603 to 325470) Timms,K.M., Lu,F., Shen,Y., Pierson,C.A., Muzny,D.M., Gu,Y., Nelson,D.L. and Gibbs,R.A.		
TITLE	130 kb of DNA sequence reveals two new genes and a regional duplication distal to the human iduronate-2-sulfatase locus		
JOURNAL	Genome Res. 5 (1), 71-78 (1995)		
MEDLINE	96352905		
REFERENCE	2 (bases 120609 to 326663)		
AUTHORS	Timms,K.M., Bondeson,M.L., Ansari-Lari,M.A., Lagerstedt,K., Muzny,D.M., Dugan-Rocha,S.P., Nelson,D.L., Pettersson,U. and Gibbs,R.A.		
TITLE	Molecular and phenotypic variation in patients with severe Hunter syndrome		
JOURNAL	Hum. Mol. Genet. 6 (3), 479-486 (1997)		
MEDLINE	97227295		

```
/rpt_family="THR"
complement(19772..20114)
repeat_region
/rpt_family="THE1b"
26241..26504
repeat_region
/rpt_family="MER2"
27126..27410
repeat_region
/rpt_family="Alu"
complement(27721..28090)
repeat_region
/rpt_family="MLT1b"
complement(29141..29221)
repeat_region
/rpt_family="MIR"
complement(29971..30214)
STS
/db_xref=dbSTS:G04466"
complement(31323..32192)
repeat_region
/rpt_family="L1"
32202..33477
repeat_region
/rpt_family="L1"
33649..33828
repeat_region
/rpt_family="MLT1f"
complement(34208..34640)
STS
/db_xref=dbSTS:L18338"
complement(37847..37896)
repeat_region
/rpt_family="MIR"
complement(37897..38047)
repeat_region
/rpt_family="THE1b"
39003..39413
repeat_region
/rpt_family="Alu"
39447..39616
repeat_region
/rpt_family="Alu"
complement(42772..42890)
repeat_region
/rpt_family="MIR2"
43946..44051
repeat_region
/rpt_family="L1PA7"
44392..44469
repeat_region
/rpt_family="MIR2"
complement(44547..44844)
repeat_region
/rpt_family="L1"
complement(44932..45060)
repeat_region
/rpt_family="L1PA15"
complement(45326..45416)
repeat_region
/rpt_family="MIR2"
45428..45935
repeat_region
/rpt_family="MLT2B2"
47122..47237
repeat_region
/rpt_family="MIR"
complement(49106..49258)
repeat_region
/rpt_family="Alu"
complement(56772..56846)
repeat_region
/rpt_family="Alu"
57858..58215
repeat_region
/rpt_family="MLT1a"
complement(59322..58680)
repeat_region
/rpt_family="MLT1a"
complement(58693..58780)
repeat_region
/rpt_family="MLT1a"
58884..59162
repeat_region
/rpt_family="Alu"
complement(59176..59299)
repeat_region
/rpt_family="MSTc"
complement(61342..61403)
repeat_region
/rpt_family="MIR"
63434..63486
repeat_region
/rpt_family="MIR"
complement(64072..64357)
repeat_region
/rpt_family="Alu"
64390..64546
repeat_region
/rpt_family="MIR"
complement(65193..65451)
STS
/db_xref=dbSTS:Z23490"
65883..66218
repeat_region
/rpt_family="MSTa"
68209..68312
repeat_region
/rpt_family="MIR"

59326..69431
/rpt_family="MLT2C2"
complement(69641..69770)
STS
/db_xref=dbSTS:L39873"
73923..74072
repeat_region
/rpt_family="MER42c"
75581..75674
repeat_region
/rpt_family="MIR"
complement(77945..78229)
repeat_region
/rpt_family="LTR1"
complement(81874..81938)
repeat_region
/rpt_family="MIR"
82148..82576
repeat_region
/rpt_family="LTR7"
86925..86957
repeat_region
/rpt_family="LTR7"
87860..88147
repeat_region
/rpt_family="MER4B"
90287..90536
repeat_region
/rpt_family="MLT1e"
92315..92507
repeat_region
/rpt_family="MIR"
complement(93107..93413)
misc_feature
/note="similar to EST with GenBank Accession Number
Z33565"
repeat_region
93873..94223
/rpt_family="MLT1a"
repeat_region
complement(94579..94636)

...: remainder of annotations omitted.

Query Match 43.3%; Score 995; DB 9; Length 326663;
Best Local Similarity 100.0%; Pred. No. 0.00e+00;
Matches 995; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Db 248953 GAAAAGAAACATGGATTACATGTTTAAACAATAATATTTTGAATATATATAGGCTCATAGAT 249012
Cp 2296 GAAAAGAAACATGGATTACATGTTTAAACAATAATATTTTGAATATATATAGGCTCATAGAT 2237
Db 249013 ATATTATTAAATATATAACAAAATAATCAGGAACCTAGTGTGGTGCCTAGTTTGATATAT 249072
Cp 2236 ATATTATTAAATATATAACAAAATAATCAGGAACCTAGTGTGGTGCCTAGTTTGATATAT 2177
Db 249073 GATTACTTTTGAATGTCACATAAATCCACAAATAATGAAAGTATTCCTTTGTGTAATG 249132
Cp 2176 GATTACTTTTGAATGTCACATAAATCCACAAATAATGAAAGTATTCCTTTGTGTAATG 2117
Db 249133 TTATGTTTGGTTATTATGATGTTGCTTCGTATCCAAAGGTATGACATAAATTCAGTTTGT 249192
Cp 2116 TTATGTTTGGTTATTATGATGTTGCTTCGTATCCAAAGGTATGACATAAATTCAGTTTGT 2057
Db 249193 TTGCTTTGTATTATTTCAGTAATAAAGCCGTAAGCTTTTAAAAAGAGGAAATTTAAAA 249252
Cp 2056 TTGCTTTGTATTATTTCAGTAATAAAGCCGTAAGCTTTTAAAAAGAGGAAATTTAAAA 1997
Db 249253 AAAAAGTGTCCAATTACCAATTATAAATTTTAAATAAGACTAAACGAAAGGTTTGGCT 249312
Cp 1996 AAAAAGTGTCCAATTACCAATTATAAATTTTAAATAAGACTAAACGAAAGGTTTGGCT 1937
Db 249313 GTTACATATTCTCAGGCCAAATTTGTTGATGTCATGTTTGGATTAACTAGCCCTCAGCTGC 249372
Cp 1936 GTTACATATTCTCAGGCCAAATTTGTTGATGTCATGTTTGGATTAACTAGCCCTCAGCTGC 1877
Db 249373 TTCCAATATTATGGGTAATFACAAAACAGCAGCTCTAACTCCTCCTCACCAGCTGGA 249432
Cp 1876 TTCCAATATTATGGGTAATFACAAAACAGCAGCTCTAACTCCTCCTCACCAGCTGGA 1817
Db 249433 AGGGAGCACATCACATTTGCCATCCATGTTGGCAAAACTCAAGGCATCAACACTGGAA 249492
Cp 1816 AGGGAGCACATCACATTTGCCATCCATGTTGGCAAAACTCAAGGCATCAACACTGGAA 1757
Db 249493 AAGATCTCCACCTTGGGAATCATTTATACATATTGTGATCCTCAATGGGTGAGATCCAC 249552
|||||
```



BASE COUNT	494 a	375 c	444 g	518 t	
ORIGIN					
Query Match	18.9%;	Score 434;	DB 10;	Length 1831;	
Best Local Similarity	85.0%;	Pred. No. 0.00e+00;			
Matches	533;	Conservative	0;	Mismatches 93;	Indels 1;
Gaps	1;				
Db	1	CTTCTGGCAGTTGGGTACCAACAGCCGCCATATGCCCTTCAGATACCCAGGAATTTCA	60		
Qy	781	CTTCTGGCCCTGGGTATAGCCACACATCCCTTCAGATACCCAGGAATTTCA	840		
Db	61	GAAGTTGCATCCCTTGGAAACATACACCTGCTCTCATCCCATGTTCTCTGATAGCCT	120		
Qy	841	GAAGTTGTATCCCTTGGAAACATACACCTGCTCTCATCCCATGTTCTCTGATAGCCT	900		
Db	121	ACCACCTGTAGCTACACCCCTGGATGATATCAGGAGAGGAAGATGTCCAAAGCCTT	180		
Qy	901	ACCACCTGTAGCTACACCCCTGGATGATATCAGGAGAGGAAGATGTCCAAAGCCTT	960		
Db	181	AAACATCAGTGTCCCTATGAGCAACATTCCTGAGGATTTTCAGCGGAAATCCCGCCAG	240		
Qy	961	AAACATCAGTGTCCCTATGAGTGTCCAAATTCCTGTGACTTTTCAGCGGAAATCCCGCCAG	1020		
Db	241	CTACTTTGCTTCTGTGCTATTTGGATACATCAGTCGAGCATGTTTGAAGTCTTTGA	300		
Qy	1021	CTACTTTGCTTCTGTGCTATTTGGATACATCAGTCGAGCATGTTTGAAGTCTTTGA	1080		
Db	301	TGATCTTCTGTAGCCCAACACAAATATTGCTTTTCATCTGATCAGGATGGGCCCT	360		
Qy	1081	CGATCTTCTGTAGCCCAACACCAATATTGCTTTTCATCTGATCAGGATGGGCCCT	1140		
Db	361	AGGTGAACACGGAGATGGGCAAGTACAGCAACATTTGATGTGCTACCCCTGTGCCCT	420		
Qy	1141	AGGTGAACATGGAGATGGGCAAGTACAGCAATTTGATGTGCTACCCCTGTGCCCT	1200		
Db	421	GATGCTATGCTCCTGGAAGAGCGGCCCACTTCCTCGGAGGCGCCAGAGCTCTTCC	480		
Qy	1201	GATATTTCTATGCTTCTGGAAGAGCGGCCCTTCTCCTCGGAGGCGCCAGAGCTCTTCC	1260		
Db	481	ATACCGGAGCCCTTTGATCCGGCTCGGATGATGATGATGATGATGATGATGATGAT	540		
Qy	1261	TTACTTCGACCTTTTGAATCCGGCTCACAGTTGATGATGATGATGATGATGATGAT	1320		
Db	541	CCTGTGAACTTGTATCTCTTCCCACTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT	599		
Qy	1321	CCTGTGAACTTGTATCTCTTCCCACTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT	1380		
Db	600	TCCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT	626		
Qy	1381	ACCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT	1407		
RESULT	13	GI3552	400 bp	DNA	STS
LOCUS	human STS SHGC-11149.				04-JAN-1996
DEFINITION	human STS SHGC-11149.				
ACCESSION	GI3552				
VERSION	GI3552.1	GI:1129291			
KEYWORDS	STS sequence; primer; sequence tagged site.				
SOURCE	human.				
ORGANISM	Homo sapiens				
	Eukaryotae; mitochondrial eukaryotes; Metazoa; Chordata;				
	Vertebrata; Gnathostomata; Osteichthyes; Sarcopterygii; Chonata;				
	Tetrapoda; Amniota; Mammalia; Theria; Eutheria; Archonta; Primates;				
	Catarrhini; Hominoidea; Homo.				
	1 (bases 1 to 400)				
	Myers, R.M.				
	Unpublished (1995)				
REFERENCE					
AUTHORS	Contact: Richard M. Myers				
JOURNAL	Stanford Human Genome Center (SHGC)				
COMMENT	Stanford University School of Medicine				
	Department of Genetics, M-344, Stanford, CA 94305, USA				
	Tel: 4157259687				

Fax: 4157259689					
Email: myers@hgc.stanford.edu					
Primer A: TGGTGAAGGAGGAGTTCAGAGC					
Primer B: GGCTGTTACATATCTCAGGCC					
STS size: 118					
PCR Profile:					
Initial incubation: 94 degrees C for 90 seconds					
Denaturation: 94 degrees C for 15 seconds					
Annealing: 62 degrees C for 23 seconds					
Polymerization: 72 degrees C for 30 seconds					
PCR Cycles: 30					
Thermal Cycler: Perkin Elmer 9600					
Protocol:					
Template: 25 ng					
Primer: each 1 uM					
dNTPs: each 200 uM					
Taq Polymerase: 0.05 units/uL					
Total Vol: 10 uL					
Buffer:					
MgCl2: 2.5 mM					
KCl: 50 mM					
Tris-HCl: 20 mM					
pH: 8.3					
Prepared with primer pairs derived from M58342 -- Unigene.					
Location/Qualifiers					
1..400					
/organism="Homo sapiens"					
STS					
46..163					
primer_bind					
46..67					
primer_bind					
complement(142..163)					
BASE COUNT	132 a	69 c	68 g	121 t	10 others
ORIGIN					
Query Match	17.0%;	Score 390;	DB 11;	Length 400;	
Best Local Similarity	97.5%;	Pred. No. 1.49e-290;			
Matches	390;	Conservative	0;	Mismatches 10;	Indels 0;
Gaps	0;				
Db	1	GTTTGGCCCAACCATGGATGGCAAAATGTGATGTCTCCCTTCCAGCTGGTGAGAGGAG	60		
Qy	1778	GTTTGGCCCAACCATGGATGGCAAAATGTGATGTCTCCCTTCCAGCTGGTGAGAGGAG	1837		
Db	61	TTAGAGCTGGTGGTTTGTGATTTACCCATTAATTTGGAAGCAGCCCTGAGGCTAGTTAA	120		
Qy	1838	TTAGAGCTGGTGGTTTGTGATTTACCCATTAATTTGGAAGCAGCCCTGAGGCTAGTTAA	1897		
Db	121	CCAAACATGCATCAACAATTTGGCCCTGAGAATATGTAACAGCCAAACCTTTTCGTTTAG	180		
Qy	1898	CCAAACATGCATCAACAATTTGGCCCTGAGAATATGTAACAGCCAAACCTTTTCGTTTAG	1957		
Db	181	CTTTTAAATTTTAAATTTGTTAAATTTGGACCAAGNNNNNNNNNNAAATTTCCCTCTTTTA	240		
Qy	1958	CTTTTAAATTTTAAATTTGTTAAATTTGGACCAAGNNNNNNNNNNAAATTTCCCTCTTTTA	2017		
Db	241	AAACAGTTACGGCTTATTACTGATAAATACAAACCAACCAAACTCAAGTTATGTCATA	300		
Qy	2018	AAACAGTTACGGCTTATTACTGATAAATACAAACCAACCAAACTCAAGTTATGTCATA	2077		
Db	301	CCCTTGGATACCAAGCAGCCATACATATTAACCAACCAATATATACACAAAGAACTT	360		
Qy	2078	CCCTTGGATACCAAGCAGCCATACATATTAACCAACCAATATATACACAAAGAACTT	2137		
Db	361	TCATTATTTGTGGAATTTAGTGCATTTCAAAAAGTAATCA	400		
Qy	2138	TCATTATTTGTGGAATTTAGTGCATTTCAAAAAGTAATCA	2177		
RESULT	14	AC002315	110079 bp	DNA	ROD
LOCUS	Mouse BAC-146N21				20-FEB-1998
DEFINITION	Mouse BAC-146N21 Chromosome x contains iduronate-2-sulfatase gene;				



## complete sequence.

AC002315  
 VERSION  
 AC002315.1 GI:2258164

KEYWORDS  
 HTG.  
 SOURCE  
 house mouse.

## ORGANISM

Mus musculus  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;  
 Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
 1 (bases 1 to 110079)  
 Timms,K.M., Hockett,L., Edwards,F.J., Lu,J., Muzny,D.M., Miller,W.  
 and Gibbs,R.A.

## TITLE

Cross-species sequence comparison of the IDS region

## JOURNAL

## REFERENCE

AUTHORS  
 Muzny,D., Ansari-Lari,M.A., Timms,K.M., Yu,W., Dugan,S., Lu,J.,  
 Shen,Y., Rowland,K., Liu,W., Perez,L., Ding,Y., Haywood,M.,  
 Jain,A., Leal,B., Logan,O., Nguyen,V., Savage,L., Shen,H.,  
 Worley,K., Chen,E., Forcum,J., Arenson,A.D., Chiu,M.W.,  
 Gorrell,J.H., Brundage,E., Di,W., Chinault,C., Nelson,D. and  
 Gibbs,R.A.

## TITLE

Direct Submission

## JOURNAL

REFERENCE  
 3 (bases 1 to 110079)

## AUTHORS

Chiu,M.W.

## TITLE

Direct Submission

## JOURNAL

Submitted (15-JUL-1997) Molecular and Human Genetics, Baylor  
 College of Medicine, One Baylor Plaza, Houston, TX 77030, USA  
 Sequencing is completed to a minimum standard of double strand  
 coverage with a minimum of 2 clones and 2 reads with no ambiguities  
 or 2 chemistries with a minimum of 2 clones and 3 reads with no  
 ambiguities. If the sequence quality does not meet this standard,  
 it will be indicated in the annotation.

## COMMENT

The repeat regions shown were identified using RepeatMasker by  
 Adrian Smit.

Sequence similarities were identified using PowerBlast by Jinghui  
 Zhang.

Exon/Intron boundaries of identified genes were chosen if there  
 were canonical splice junctions that maintained sequence continuity  
 across the splice junctions.

## FEATURES

## source

1. .110079  
 /organism="Mus musculus"  
 /db\_xref="taxon:10090"  
 /chromosome="X"  
 /clone="BAC-146N21"  
 244..561  
 /rpt\_family="LTR/MaLR"  
 567..2305  
 /rpt\_family="LTR/MaLR"  
 2306..2646  
 /rpt\_family="LTR/MaLR"  
 complement(2864..3101)  
 /rpt\_family="LTR/Retroviral"  
 complement(3812..3847)  
 /rpt\_family="Simple\_repeat"  
 complement(4982..5064)  
 /rpt\_family="Simple\_repeat"  
 complement(5065..5184)  
 /rpt\_family="LINE/L1"  
 complement(5257..5457)  
 /rpt\_family="SINE/B2"  
 6904..6953  
 /rpt\_family="Simple\_repeat"  
 complement(7980..8199)  
 /rpt\_family="SINE/B2"  
 complement(10494..10582)  
 /rpt\_family="SINE/Alu"  
 complement(10633..10861)  
 /rpt\_family="LTR/MaLR"  
 complement(10892..11185)

repeat\_region  
 complement(11193..11366)  
 /rpt\_family="LTR/MaLR"  
 11629..11718  
 /rpt\_family="Simple\_repeat"  
 complement(11839..11971)  
 /rpt\_family="SINE/B4"  
 complement(12138..12171)  
 /rpt\_family="Low\_complexity"  
 complement(12276..12396)  
 /rpt\_family="SINE/B2"  
 14061..14118  
 /rpt\_family="Simple\_repeat"  
 14096..14220  
 /note="Single strand coverage only"  
 14179..14201  
 /rpt\_family="Low\_complexity"  
 join(14201..14287,15145..15281,15929..16106,17952..18040,  
 18934..19134,20646..20816,25808..25934,27982..28155)  
 /gene="IDS"  
 complement(14424..14541)  
 /rpt\_family="Simple\_repeat"  
 complement(14783..14914)  
 /rpt\_family="Simple\_repeat"  
 17009..17084  
 /rpt\_family="SINE/MIR"  
 complement(17555..17697)  
 /rpt\_family="SINE/Alu"  
 complement(17620..17719)  
 /rpt\_family="LTR/MaLR"  
 complement(17821..17869)  
 /rpt\_family="SINE/MIR"  
 complement(18187..18314)  
 /rpt\_family="SINE/Alu"  
 18371..18499  
 /rpt\_family="SINE/B4"  
 19306..19411  
 /rpt\_family="SINE/Alu"  
 complement(19902..20388)  
 /rpt\_family="Other/MER21\_gro"  
 complement(21582..21661)  
 /rpt\_family="Simple\_repeat"  
 complement(21662..21761)  
 /rpt\_family="scRNA"  
 complement(21684..21761)  
 /rpt\_family="SINE/ID"  
 21812..21849  
 /rpt\_family="DNA/MER1\_type"  
 complement(23849..23940)  
 /rpt\_family="SINE/Alu"  
 25152..25220  
 /rpt\_family="LTR/MaLR"  
 complement(26165..26376)  
 /rpt\_family="SINE/B2"  
 complement(26379..26577)  
 /rpt\_family="SINE/B2"  
 26995..27125  
 /rpt\_family="SINE/B2"  
 27322..27505  
 /rpt\_family="SINE/B2"  
 complement(27695..27734)  
 29292..30131  
 /rpt\_family="Low\_complexity"  
 /rpt\_family="LINE/L1"  
 30204..30233  
 /rpt\_family="Simple\_repeat"  
 complement(31179..31235)  
 /rpt\_family="Simple\_repeat"  
 31236..31266  
 /rpt\_family="Simple\_repeat"  
 31579..31703  
 /rpt\_family="Simple\_repeat"  
 31705..32083



\*\*\*\*\*

W P S R E H

(TM)

\*\*\*\*\*

Release 3.1A John F. Collins, Biocomputing Research Unit.  
Copyright (c) 1993-1998 University of Edinburgh, U.K.  
Distribution rights by Oxford Molecular Ltd

MPsrch\_pp protein - protein database search, using Smith-Waterman algorithm

Run on: Wed Apr 19 21:16:13 2000; MasPar time 44.92 Seconds  
Tabular output not generated. 848.838 Million cell updates/sec

Title: >US-09-249-003-2  
Description: (1-550) from US09249003.pep  
Perfect Score: 4069  
Sequence: 1 MPPPTGRGLLGLVLSV.....QDHMYNDGGLFQLLMP 550

Scoring table: PAM 150  
Gap 11

Searched: 225878 seqs, 69334122 residues

Post-processing: Minimum Match 0%  
Listing first 45 summaries

Database: sptrmb112

1:sp\_archaea 2:sp\_bacteria 3:sp\_fungi 4:sp\_human  
5:sp\_invertebrate 6:sp\_mammal 7:sp\_plant 8:sp\_organelle  
9:sp\_phage 10:sp\_plant 11:sp\_rodent 12:sp\_unclassified  
13:sp\_vertebrate 14:sp\_virus

Statistics: Mean 50.386; Variance 92.586; scale 0.544

Pred. No. is the number of results predicted by chance to have a  
score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description	Pred. No.
1	2445	60.1	343	4	IDURONATE-2-SULPHATASE	0.00e+00
2	1036	25.5	179	4	IDURONATE-2-SULPHATASE	1.00e-194
3	604	14.8	86	4	IDURONATE-2-SULPHATASE	8.45e-99
4	338	8.3	512	2	CHOLINE SULFATASE (EC	1.46e-42
5	190	4.7	551	5	ARYLSULFATASE	3.92e-14
6	185	4.5	514	2	PHOSPHONATE MONOESTER	2.97e-13
7	182	4.5	970	2	HYPOTHETICAL 105.7 KD	9.93e-13
8	179	4.4	465	2	ARYSULFATASE	3.29e-12
9	174	4.3	559	5	ARYLSULFATASE	2.39e-11
10	143	3.5	452	5	SIMILAR TO ARYLSULFATA	3.09e-06
11	134	3.3	525	4	KIAA1001 PROTEIN.	7.70e-05
12	136	3.3	787	2	ARYLSULFATASE	3.80e-05
13	133	3.3	787	2	ARYLSULFATASE	3.80e-05
14	126	3.1	571	2	HYPOTHETICAL 86.1 KD P	1.09e-04
15	123	3.0	577	2	HYPOTHETICAL 64.1 KD P	1.22e-03
16	121	3.0	580	2	ARYLSULFATASE PRECURSOR	3.35e-03
17	117	2.9	609	3	POLAR FLAGELLAR M-RING	6.51e-03
18	112	2.8	495	2	HYPOTHETICAL C2H2-TYPE	2.41e-02
19	108	2.7	371	8	PUTATIVE 54.5 KDA PROT	1.19e-01
20	108	2.7	374	2	CYTOCHROME OXIDASE I (	4.12e-01
					OMP37 PORIN PRECURSOR	4.12e-01

21	111	2.7	409	8	Q9XP15	CYTOCHROME C OXIDASE S	1.63e-01
22	111	2.7	708	1	Q9YC40	708AA LONG HYPOTHETICA	1.63e-01
23	109	2.7	709	5	Q21376	SIMILAR TO SULFATASE.	3.03e-01
24	108	2.7	1811	5	Q9X2U8	LARGEST SUBUNIT OF THE	4.12e-01
25	109	2.7	1889	5	Q9X2U9	LARGEST SUBUNIT OF THE	3.03e-01
26	109	2.7	1889	5	Q9X2S2	LARGEST SUBUNIT OF THE	3.03e-01
27	108	2.7	1889	5	Q9X2S1	LARGEST SUBUNIT OF THE	4.12e-01
28	111	2.7	7576	2	Q9ZGA4	FK506 POLYKETIDE SYNTH	1.63e-01
29	104	2.6	159	2	Q923T1	HALOALKANOIC ACID DEHA	1.38e+00
30	105	2.6	403	8	Q9XP03	CYTOCHROME C OXIDASE S	1.02e+00
31	107	2.6	409	8	Q9XP02	CYTOCHROME C OXIDASE S	5.59e-01
32	107	2.6	410	8	Q9XP10	CYTOCHROME C OXIDASE S	5.59e-01
33	105	2.6	647	2	O84272	TRANSGLYCOLASE/TRANSP	1.02e+00
34	106	2.6	667	2	P71538	PROBABLE CARBAMOYL-PHO	7.58e-01
35	106	2.6	815	5	Q18385	SIMILAR TO PROTEIN TYR	7.58e-01
36	107	2.6	1034	11	Q61856	ORF.	5.59e-01
37	107	2.6	3262	4	Q13788	APOLIPOPROTEIN B-100 (	5.59e-01
38	101	2.5	280	2	P71422	MDCB.	3.34e+00
39	102	2.5	366	14	O86317	NON-STRUCTURAL SIGMA-N	2.49e+00
40	101	2.5	366	14	O86308	NON-STRUCTURAL SIGMA-N	3.34e+00
41	103	2.5	409	8	Q9XP12	CYTOCHROME C OXIDASE S	1.86e+00
42	103	2.5	410	8	Q9XP16	CYTOCHROME C OXIDASE S	1.86e+00
43	103	2.5	440	2	P75608	R02_ORF440 PROTEIN.	1.86e+00
44	103	2.5	467	2	O06256	HYPOTHETICAL 47.6 KD P	1.86e+00
45	103	2.5	518	2	O68838	GAMMA-GLUTAMYL-CYSTEINE	1.86e+00

ALIGNMENTS

RESULT	ID	Q14604	PRELIMINARY;	PRT;	343 AA.
AC	Q14604;				
DT	01-NOV-1996 (TREMBLrel. 01, Created)				
DT	01-NOV-1996 (TREMBLrel. 01, Last sequence update)				
DT	01-NOV-1996 (TREMBLrel. 12, Last annotation update)				
DE	IDURONATE-2-SULPHATASE.				
GN	IDS.				
OS	Homo sapiens (Human).				
OC	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;				
OC	Eutheria; Primates; Catarrhini; Hominidae; Homo.				
RN	[1]				
RP	SEQUENCE FROM N.A.				
RC	TISSUE-LYMPHOCTE;				
RX	MEDLINE; 96079126.				
RA	MALMGREN H., CARLBERG B.M., PETERSSON U., BONDESON M.L.;				
RT	"Identification of an alternative transcript from the human				
RT	iduronate-2-sulfatase (IDS) gene.";				
RL	Genomics 29:291-293(1995).				
DR	EMBL; L40586; AAA92014.1; -				
DR	PROSITE; PS00149; SULFATASE_2; 1.				
DR	PROSITE; PS00523; SULFATASE_1; 1.				
DR	PFAM; PF00884; Sulfatase; 1.				
SQ	SEQUENCE 343 AA; 38310 MW; F1825B26 CRC32;				

Query Match 60.1%; Score 2445; DB 4; Length 343;  
Best Local Similarity 99.1%; Pred. No. 0.00e+00;  
Matches 336; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

Db	1	MPPPTGRGLLGLVLSV	CVVCSVALGSETQANSTTDALNVLLIIIVDDLRPSICGCGDKLVR	60
QY	1	MPPPTGRGLLGLVLSV	CVVCSVALGSETQANSTTDALNVLLIIIVDDLRPSICGCGDKLVR	60
Db	61	SPNIDQLASHLSLFFQNA	FAQAQVACAPSRVSFLTGRRRDPTTLRLYDFNSYWRVHAGNFSTIP	120
QY	61	SPNIDQLASHLSLFFQNA	FAQAQVACAPSRVSFLTGRRRDPTTLRLYDFNSYWRVHAGNFSTIP	120
Db	121	QYFKENGYVTMSVGKVF	HPGPGISSNHTDDSPYSFSPHPSPSEKYEKTKTCRPGDGLHA	180
QY	121	QYFKENGYVTMSVGKVF	HPGPGISSNHTDDSPYSFSPHPSPSEKYEKTKTCRPGDGLHA	180
Db	181	NLLCPVDVLDVPGTLP	DKOSTEQATQLLEKMKTSASPFPLAVGYHKPHIPFPYKPEFOK	240
QY	181	NLLCPVDVLDVPGTLP	DKOSTEQATQLLEKMKTSASPFPLAVGYHKPHIPFPYKPEFOK	240



RC	STRAIN=PG2982;
RX	MEDLINE; 96421555.
RA	DOTSON S.B., SMITH C.E., LING C.S., BARRY G.F., KISHORE G.M.;
RB	"Identification, characterization, and cloning of a phosphonate
RT	monoster hydrolase from Burkholderia caryophylli PG2982.";
RL	J. Biol. Chem. 271:25754-25761(1996).
RR	EMBL; U44852; AAC44467.1; -.
DR	PFAM; PF00884; Sulfatase; 1.
KW	Hydrolase.
SQ	SEQUENCE 514 AA; 58176 MW; AF075B60 CRC32;
Query Match 4.5%; Score 185; DB 2; Length 514;	
Best Local Similarity 28.0%; Pred. No. 2.97e-13;	
Matches 46; Conservative 47; Mismatches 62; Indels 9; Gaps 9;	
Db	193 TERALTYL-TGR-DCKPFFHLGYRPHFPFVASPYHAMKYADMPAPTRAENPDA-EA 249    :    :   :     :     :     :     :     :     :     :     :
Qy	202 TEQAQLLEKMKTSASPFLLAVGVHKPHIPRPYKFQKLYPLENTIAPDPEVPGGLPP 261    :     :   :     :     :     :     :     :     :     :     :
Db	250 AQ-HPLMKHYIDHIRRGSEFHGAGSGATLDEGEIROMRATYCGLITEIDDCLGRVFAYL 308    :     :   :     :     :     :     :     :     :     :     :
Qy	262 VAYNPWM-D-IRREDVQALNISVPYGPIPVDFQ-RKIROSYPASYSYLTQTQVGRLLSAL 318    :     :   :     :     :     :     :     :     :     :     :
Db	309 DETGQWDTLIIFTSDHGEOGDHLLHGKIG-YNAESFRILPLVI 351    :     :   :     :     :     :     :     :     :     :     :
Qy	319 DDLQLANSTIAFTSDHGHWALGEHGEWAKYSNFDVAT-HVPLIF 361    :     :   :     :     :     :     :     :     :     :     :
RESULT	7
ID	O65931 PRELIMINARY; PRT; 970 AA.
AC	O65931;
DT	01-AUG-1998 (TrEMBLrel. 07, Created)
DT	01-AUG-1998 (TrEMBLrel. 07, Last sequence update)
DT	01-MAY-1999 (TrEMBLrel. 10, Last annotation update)
DE	HYPOTHETICAL 105.7 KD PROTEIN.
GN	MTC1418A.OIC.
OS	Mycobacterium tuberculosis.
OC	Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;
OC	Actinomycetales; Corynebacterineae; Mycobacteriaceae; Mycobacterium.
RN	[1]
RC	SEQUENCE FROM N.A.
RC	STRAIN=H37RV;
RA	OLIVER K., HARRIS D.;
RL	Submitted (JUN-1997) to the EMBL/GenBank/DDBJ databases.
RN	[2]
RC	SEQUENCE FROM N.A.
RC	STRAIN=H37RV;
RA	PARKHILL J., BARRELL B.G., RAJANDREAM M.A.;
RL	Submitted (JUN-1997) to the EMBL/GenBank/DDBJ databases.
RN	[3]
RC	SEQUENCE FROM N.A.
RC	STRAIN=H37RV;
RX	MEDLINE; 96181548.
RA	PHILIPP W.J., POULET S., EIGLMEIER K., PASCOPELLA L.,
RA	BALASUBRAMANIAN V., HEYM B., BERGH S., BLOOM B.R., JACOBS W.R. JR.,
RA	COLE S.T.;
RT	"An integrated map of the genome of the tubercle bacillus,
RT	Mycobacterium tuberculosis H37RV, and comparison with Mycobacterium
RT	leprae.";
RL	Proc. Natl. Acad. Sci. U.S.A. 93:3132-3137(1996).
DR	EMBL; Z96070; CAB09444.1; -.
DR	PFAM; PF00884; Sulfatase; 1.
KW	Hypothetical protein.
SQ	SEQUENCE 970 AA; 105679 MW; 8ECA7230 CRC32;
Query Match 4.5%; Score 182; DB 2; Length 970;	
Best Local Similarity 30.0%; Pred. No. 9.93e-13;	
Matches 39; Conservative 39; Mismatches 40; Indels 12; Gaps 9;	
Db	210 DARNVLIVLDLAGFGGPDTFGGA-ITPTLSLAQNGLIY-NRFHVTVACPSTRALILT 267    :     :   :     :     :     :     :     :     :     :     :
Qy	35 DALNVLLIIVDDLRL-PSLGCYGDKLVRSPNIDLASHSLLFQAFQAQVACAPSRSVFLT 93    :     :   :     :     :     :     :     :     :     :     :







RL Proc. Natl. Acad. Sci. U.S.A. 93:3132-3137(1996).

DR EMBL; 295972; CAB09383.1; .

DR PFAM; PF00884; Sulfatase; 1.

KW Hypothetical protein.

SQ SEQUENCE 787 AA; 86127 MW; 5E38F5FF CRC32;

Query Match 3.3%; Score 133; DB 2; Length 787;

Best Local Similarity 32.9%; Pred. No. 1.09e-04;

Matches 24; Conservative 22; Mismatches 24; Indels 3; Gaps 3;

Db 35 QLRPAGAPNVLLILLDDVGFAGSSAFGGP-CRTSTAELLAGNLRY-NREHTTALCSPT 92

QY 29 QANSTTDALNVLLIIVDDLRL-PSLGCYGDKLVRSPNIDQLASHLLFQNAFAQAVCAPS 87

Db 93 QOALLTCRNHSA 105

QY 88 RVSELTGRRPDTT 100

#### RESULT 14

ID P77318 PRELIMINARY; PRT; 571 AA.

AC P77318; P78159;

DT 01-FEB-1997 (TremBLrel. 02, Created)

DT 01-FEB-1997 (TremBLrel. 02, Last sequence update)

DT 01-NOV-1999 (TremBLrel. 12, Last annotation update)

DE HYPOTHETICAL 64.1 KD PROTEIN.

OS Escherichia coli.

OC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;

OC Escherichia.

RN SEQUENCE FROM N.A.

RC STRAIN=K-12;

RA BLATTNER F.R., PLUNKETT G. III, MAYHEW G.F., PERNA N.T., GLASNER F.D.;

RL Submitted (JAN-1997) to the EMBL/GenBank/DBJ databases.

RN SEQUENCE FROM N.A.

RC STRAIN=K12;

RA AIBA H., BABA T., FUJITA K., HAYASHI K., HONJO A., HORIUCHI T.,

RA IKEMOTO K., INADA T., ISONO K., ITOH T., KANAI K., KASAI H.,

RA KASHIMOTO K., KIM S., KIMURA S., KITAGAWA M., KITAKAWA M., MAKINO K.,

RA MASUDA S., MIKI T., MIZOBUCHI K., MORI H., MOTOMURA K., NAKAMURA Y.,

RA NASHIMOTO H., NISHIO Y., OSHIMA T., SAITO N., SAMPEI G., SEKI Y.,

RA TAGAMI H., TAKEKOTO K., WADA C., YAMAMOTO Y., YANO M.;

RL Submitted (DEC-1996) to the EMBL/GenBank/DBJ databases.

RN SEQUENCE OF 90-571 FROM N.A.

RC STRAIN=K12;

RA AIBA H., BABA T., FUJITA K., HAYASHI K., INADA T., ISONO K., ITOH T.,

RA KASAI H., KASHIMOTO K., KIMURA S., KITAGAWA M., KITAGAWA M.,

RA MAKINO K., MIKI T., MIZOBUCHI K., MORI H., MORI T., MOTOMURA K.,

RA NAKADE S., NAKAMURA Y., NASHIMOTO H., NISHIO Y., OSHIMA T., SAITO N.,

RA SAMPEI G., SEKI Y., SIVASUNDARAM S., TAGAMI H., TAKEKOTO K.,

RA TAKEKOTO K., TAKEUCHI Y., WADA C., YAMAMOTO Y., HORIUCHI T.;

RT "A 570-kb DNA sequence of the Escherichia coli K-12 genome

corresponding to the 28.0-40.1 min region on the linkage map.;"

RL DNA Res. 3:363-377(1996).

DR EMBL; AE000247; AAC74571.1; .

DR EMBL; D90792; BAA15172.1; .

DR EMBL; D90791; BAA15169.1; .

DR PFAM; PF00884; Sulfatase; 1.

KW Hydrolase.

SQ SEQUENCE 571 AA; 64094 MW; 447178C5 CRC32;

Query Match 3.1%; Score 126; DB 2; Length 571;

Best Local Similarity 25.0%; Pred. No. 1.22e-03;

Matches 49; Conservative 52; Mismatches 78; Indels 17; Gaps 16;

Db 120 TPTLLSMDGVRFTNGYVAHGVSGPSRAALMTGRAPARFGVYNTDAQDQIGPLTE-TFL 178

QY 61 SPNIDQLASHLLLFQNAFAQAVCAPSRLTGRRPDTTRLY-DFNSYWRVHAGNFSTI 119

Db 179 PELFQNHGYTAAVGK-WHLISKISNPVPEDKQTRDYHDNETTFSAEWQPNRGFDYFM 237

QY 120 POYFKENGIVTMSGVKVFHPG-ISSNHTDDSPYSWSFPP-YHP-SSEKYE-NYTKT--C-R 172

Db 238 GFHAAGTAYINSPSLFKNRERVPAKYISD-QLTDEAIGVVDRAKTLDOFPFMYLAYNAP 296

QY 173 GPDGELHANLLCP-V-DVLD-VPE-GTLPDKQSTEQAIQLEKMKTSASPFPLAVGYHKP 228

Db 297 HLPNDNPADQYQKQF 312

QY 229 HIPFRYPK-E-FOKLY 242

#### RESULT 15

ID Q9X759 PRELIMINARY; PRT; 577 AA.

AC Q9X759;

DT 01-NOV-1999 (TremBLrel. 12, Created)

DT 01-NOV-1999 (TremBLrel. 12, Last sequence update)

DT 01-NOV-1999 (TremBLrel. 12, Last annotation update)

DE ARYLSULFATASE PRECURSOR (EC 3.1.6.1).

GN ATSA.

OS Klebsiella pneumoniae.

OC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;

OC Klebsiella.

RN [1]

RP SEQUENCE FROM N.A.

RC STRAIN=DSM 681;

RX MEDLINE; 99269066.

RA SZAMET C., MIECH C., BALLEININGER M., SCHMIDT B., VON FIGURA K.,

RA DIERKS T.;

RT "The iron sulfur protein AtsB is required for posttranslational

formation of formylglycine in the Klebsiella sulfatase.;"

RL J. Biol. Chem. 274:15375-15381(1999).

DR EMBL; AJ131525; CAB40961.1; .

DR PROSITE; PS00149; SULFATASE\_2; 1.

DR PROSITE; PS00523; SULFATASE\_1; 1.

KW Signal; Hydrolase.

FT SIGNAL 1 20 POTENTIAL.

SQ SEQUENCE 577 AA; 64155 MW; C1ECC635 CRC32;

Query Match 3.0%; Score 123; DB 2; Length 577;

Best Local Similarity 18.8%; Pred. No. 3.35e-03;

Matches 63; Conservative 107; Mismatches 138; Indels 27; Gaps 21;

Db 8 KAVSMILAGGAHA-AQQRPNVIVIIADDMGYSIDSPFGGE-IPTPNLCAMAEQGRMSQ 65

QY 18 SSVCVALGSETQANSTTDALNVLLIIVDDLRL-PSLGCYGDKLVRSPNIDQLASHLLFQ 76

Db 66 YTS-PMSAPARSMLLTGNQQAGMGWYDSTIGKEGYELRL-TDRVTTMAERFKDA 123

QY 77 AFAQAVCAPSRVSELTGRR-PDTTR-L-YD--F--NSY-WRVHAGNFSTIPOYFKEN 126

Db 124 GYNTLMAGK-WHLGVPVGPATPKERGFNHAFAPMGGTSHFNDAIP-LGTVEAFHYTYTRD 181

QY 127 GYVTSVGVKVFHPG-ISSNHTDDSPYSWSFPPHPSPSEKYENTKTCRGDGLHANLLCP 185

Db 182 GERSVLPDDFVSSEAYARQMNSWI-KATPKEQPVPAWLAFTAPHDPLQAPDEWIKRFGQ 240

QY 186 VDVLDPPEGTLPDKQSTEQAIQLEKMKTSASPFPLAVGYHKRPHFPFRYPKFEOKLYPLE 245

Db 241 YEQGYAEV-YRORIARL--KA-LGIHDDTLPPLHLELDKEWALTPQCKYKAKYMOVVA 296

QY 246 NITLAPDPEVDPGLPPVAYNPWMDIRQRED-VQALNISVPYGPPI-PVD--FORKINQSYF 301

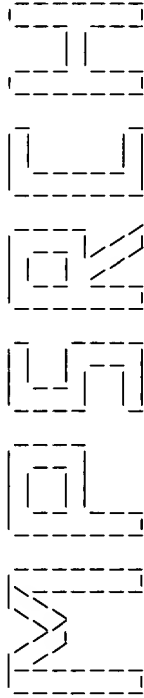
Db 297 AMIANMDAQIGTLMETLKTQGRDKNTLLVFLTDNG 331

QY 302 ASVSYLDTVQVGRLLSALDDQLANSTIIFTSDHG 336

Search completed: Wed Apr 19 21:34:11 2000

Job time : 1078 secs.

\*\*\*\*\*



Release 3.1A John F. Collins, Biocomputing Research Unit.  
Copyright (c) 1993-1998 University of Edinburgh, U.K.  
Distribution rights by Oxford Molecular Ltd

Mprsch\_pp protein - protein database search, using Smith-Waterman algorithm

Run on: Wed Apr 19 21:09:19 2000; MasPar time 19.58 Seconds  
Tabular output not generated. 838.906 Million cell updates/sec

Title: >US-09-249-003-2  
Description: (1-550) from US09249003.pap  
Perfect Score: 4069  
Sequence: 1 MPPPTGRGLLWGLVLSSV.....QDHNNYNDSSGGDLFQLLMP 550

Scoring table: PAM 150  
Gap 11

Searched: 82229 seqs, 29864866 residues

Post-processing: Minimum Match 0%  
Listing first 45 summaries

Database: swiss-prot38  
1:swissprot

Statistics: Mean 52.053; Variance 88.452; scale 0.588

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description	Pred. No.
1	4069	100.0	550	1	IDS_HUMAN	0.00e+00
2	3129	76.9	563	1	IDURONATE 2-SULFATASE	0.00e+00
3	282	6.9	497	1	YIDI_ECOLI	6.79e-34
4	266	6.5	624	1	STSL_MOUSE	1.50e-30
5	258	6.3	591	1	ARYLSULFATASE F PRECUR	6.75e-29
6	239	5.9	589	1	ARYLSULFATASE E PRECUR	5.12e-25
7	234	5.8	593	1	ARYLSULFATASE D PRECUR	5.22e-24
8	211	5.2	522	1	GA6S_HUMAN	1.89e-19
9	211	5.2	567	1	ARYLSULFATASE PRECURSO	1.89e-19
10	213	5.2	583	1	STSL_HUMAN	7.70e-20
11	202	5.0	577	1	STSL_RAT	1.05e-17
12	198	4.9	506	1	ARYLSULFATASE A PRECUR	6.17e-17
13	195	4.8	507	1	ARYLSULFATASE B PRECUR	2.04e-15
14	190	4.7	533	1	ARYLSULFATASE A PRECUR	2.04e-15
15	184	4.5	551	1	ARYLSULFATASE PRECURSO	2.73e-14
16	176	4.3	502	1	SPHM_HUMAN	8.42e-12
17	172	4.2	535	1	ARYLSULFATASE B PRECUR	4.42e-12
18	153	3.8	313	1	RIBF_ECOLI	1.04e-08
19	153	3.8	551	1	ARYLSULFATASE (EC 3.1.	1.04e-08
20	126	3.1	433	1	TCOL_HUMAN	2.92e-04
21	126	3.1	919	1	ARYLSULFATASE PRECURSO	2.92e-04
22	124	3.0	464	1	ARYLSULFATASE PRECURSO	5.95e-04
23	118	2.9	189	1	COAT_CCMV	4.83e-03

24	117	2.9	532	1	ARS_PSEAE	ARYLSULFATASE (EC 3.1.	6.80e-03
25	118	2.9	546	1	GCVK_HSV7J	POSSIBLE GANCICLOVIR K	4.83e-03
26	115	2.8	473	1	ARSB_RAT	ARYLSULFATASE B (EC 3.	1.34e-02
27	114	2.8	540	1	NADB_ECOLI	L-ASPARTATE OXIDASE (E	1.87e-02
28	114	2.8	559	1	GL6S_CAPHI	N-ACETYLGLUCOSAMINE-6-	1.87e-02
29	112	2.8	585	1	YEJM_RAEIN	HYPOTHETICAL PROTEIN H	3.65e-02
30	110	2.7	846	1	YDSB_SCHPO	HYPOTHETICAL 93.2 KD T	7.03e-02
31	105	2.6	285	1	ARSB_MOUSE	ARYLSULFATASE B (EC 3.	3.49e-01
32	107	2.6	552	1	GL6S_HUMAN	N-ACETYLGLUCOSAMINE-6-	1.85e-01
33	107	2.6	4563	1	APB_HUMAN	APOLIPOPROTEIN B-100 P	1.85e-01
34	101	2.5	190	1	YEJM_SALTY	HYPOTHETICAL PROTEIN I	1.20e+00
35	101	2.5	353	1	REBU_SALTY	REBU PROTEIN.	1.20e+00
36	101	2.5	383	1	OMS2_SALTY	OUTER MEMBRANE PROTEIN	1.20e+00
37	101	2.5	394	1	OMSL_SALTY	OUTER MEMBRANE PROTEIN	1.20e+00
38	101	2.5	510	1	GSHI_PSESM	2,3-BISPHOSPHOGLYCERAT	1.20e+00
39	102	2.5	518	1	GSHI_ECOLI	GLUTAMATE--CYSTEINE LI	8.85e-01
40	102	2.5	586	1	YEJM_ECOLI	HYPOTHETICAL 67.3 KD P	8.85e-01
41	100	2.5	1013	1	EPAS_CHICK	EPHRIN TYPE-A RECEPTOR	1.62e+00
42	100	2.5	1896	1	RPB1_DROME	DNA-DIRECTED RNA POLYM	1.62e+00
43	98	2.4	365	1	RRPO_REOVD	RNA-DIRECTED RNA POLYM	2.95e+00
44	99	2.4	377	1	OMP_N_ECOLI	OUTER MEMBRANE PROTEIN	2.19e+00
45	99	2.4	2103	1	RRPL_UUK	RNA POLYMERASE (EC 2.7	2.19e+00

ALIGNMENTS

RESULT 1  
ID IDS\_HUMAN STANDARD; PRT; 550 AA.  
AC P22304;  
DT 01-AUG-1991 (Rel. 19, Created)  
DT 01-AUG-1991 (Rel. 19, Last sequence update)  
DT 15-JUL-1999 (Rel. 38, Last annotation update)  
DE IDURONATE 2-SULFATASE PRECURSOR (EC 3.1.6.13).  
GS IDS.  
GN Homo sapiens (Human).  
OC Eukaryota; Metazoa; Chordata; Vertebrata; Mammalia;  
OC Eutheria; Primates; Catarrhini; Homnidae; Homo.  
RN [1]  
RC SEQUENCE FROM N.A., AND SEQUENCE OF 34-58 AND 456-473  
RX MEDLINE; 91046030.  
RA WILSON P.J., MORRIS C.P., ANSON D.S., OCCHIODORO T., BIELICKI J.,  
CLEMENTS P.R., HOPWOOD J.J.;  
RT "Hunter syndrome: isolation of an iduronate-2-sulfatase cDNA clone  
and analysis of patient DNA.";  
RL Proc. Natl. Acad. Sci. U.S.A. 87:8531-8535(1990).  
RN [2]  
RP SEQUENCE FROM N.A.  
RX MEDLINE; 94063929.  
RA WILSON P.J., MEANEY C.A., HOPWOOD J.J., MORRIS C.P.;  
RL "Sequence of the human iduronate 2-sulfatase (IDS) gene.";  
RN [3]  
RP SEQUENCE FROM N.A.  
RX MEDLINE; 96352905.  
RA TIMMS K.M., LU F., SHEN Y., PIERSON C.A., MUZYNY D.M., GU Y.,  
NELSON D.L., GIBBS R.A.;  
RT "130 kb of DNA sequence reveals two new genes and a regional  
duplication distal to the human iduronate-2-sulfate sulfatase  
locus.";  
RL Genome Res. 5:71-78(1995).  
RN [4]  
RP SEQUENCE FROM N.A.  
RA MUZYNY D., ANSARI-LARI M.A., TIMMS K.M., YU W., DUGAN S., LU J.,  
SHEN Y., ROWLAND K., LIU W., PEREZ L., DING Y., GONZALEZ O.,  
HAYWOOD M., JAIN A., LEAL B., LOGAN O., NGUYEN V., SAVAGE L., SHEN H.,  
WORLEY K., CHEN E., FORCUM J., ARENSON A.D., CHIU M.W., GORRELL J.H.,  
BRUNDAGE E., DI W., CHINAULT C., NELSON D., GIBBS R.A.;  
RL Submitted (AUG-1997) to the EMBL/GenBank/DBJ databases.  
RN [5]  
RP SEQUENCE OF 1-398 FROM N.A.  
RX MEDLINE; 93258349.  
RA FLOMEN R.H., GREEN E.P., GREEN P.M., BENTLEY D.R., GIANNELLI F.;

Applicant

RT "determination of the organisation of coding sequences within the  
 RT iduronate sulphate sulphatase (IDS) gene.";  
 RL Hum. Mol. Genet. 2:3-10(1993).  
 RN [6]  
 RN REVIEW ON MPS-II VARIANTS.  
 RX MEDLINE; 94154729.  
 RA HOPWOOD J.J., BUNGE S., MORRIS C.P., WILSON P.J., STEGLICH C.,  
 RA BECK M., SCHWINGER E., GAL A.;  
 RA "Molecular basis of mucopolysaccharidosis type II: mutations in the  
 RT iduronate-2-sulphatase gene.";  
 RL Hum. Mutat. 2:435-442(1993).  
 RN [7]  
 RN VARIANTS MPS-II ARG-135 AND GLY-422.  
 RX MEDLINE; 9325059.  
 RA BUNGE S., STEGLICH C., BECK M., ROSENKRANZ W., SCHWINGER E.,  
 RA HOPWOOD J.J., GAL A.;  
 RA "Mutation analysis of the iduronate-2-sulphatase gene in patients with  
 RT mucopolysaccharidosis type II (Hunter syndrome).";  
 RL Hum. Mol. Genet. 1:335-339(1992).  
 RN [8]  
 RN VARIANT MPS-II TRP-468.  
 RX MEDLINE; 93258320.  
 RA CROTTI P.L., BUNGE S., ANDERSON R.A., WHITLEY C.B.;  
 RA "Mutation R468W of the iduronate-2-sulphatase gene in mild Hunter  
 RT syndrome (mucopolysaccharidosis type II) confirmed by in vitro  
 RT mutagenesis and expression.";  
 RL Hum. Mol. Genet. 1:755-757(1992).  
 RN [9]  
 RN VARIANTS MPS-II ARG-86; ASP-94; ARG-120; PRO-221 AND GLY-422.  
 RX MEDLINE; 94108441.  
 RA BUNGE S., STEGLICH C., ZUTHER C., BECK M., MORRIS C.P., SCHWINGER E.,  
 RA SCHINZEL A., HOPWOOD J.J., GAL A.;  
 RA "Iduronate-2-sulphatase gene mutations in 16 patients with  
 RT mucopolysaccharidosis type II (Hunter syndrome).";  
 RL Hum. Mol. Genet. 2:1871-1875(1993).  
 RN [10]  
 RN VARIANTS MPS-II GLU-68; HIS-293; GLY-478 AND ARG-485.  
 RX MEDLINE; 95072615.  
 RA SCHROEDER W., WULF K., WEHNERT M., SEIDLITZ G., HERRMANN F.H.;  
 RA "Mutations of the iduronate-2-sulphatase (IDS) gene in patients with  
 RT Hunter syndrome (mucopolysaccharidosis II).";  
 RL Hum. Mutat. 4:128-131(1994).  
 RN [11]  
 RN VARIANTS MPS-II TRP-132; TYR-229; ARG-358; HIS-469 AND CYS-523.  
 RX MEDLINE; 95193786.  
 RA JONSSON J.J., ARONOVICH E.L., BRAUN S.E., WHITLEY C.B.;  
 RT "Molecular diagnosis of mucopolysaccharidosis type II (Hunter  
 RT syndrome) by automated sequencing and computer-assisted  
 RT interpretation: toward mutation mapping of the iduronate-2-sulphatase  
 RT gene.";  
 RL Am. J. Hum. Genet. 56:597-607(1995).  
 RN [12]  
 RN VARIANTS MPS-II LEU-86; ASN-87; PRO-92; ASN-135; LYS-345 AND TRP-468.  
 RX MEDLINE; 95245347.  
 RA POPOWSKA E., RATHMANN M., TYLKI-SZYMANSKA A., BUNGE S., STEGLICH C.,  
 RA SCHWINGER E., GAL A.;  
 RA "Mutations of the iduronate-2-sulphatase gene in 12 Polish patients  
 RT with mucopolysaccharidosis type II (Hunter syndrome).";  
 RL Hum. Mutat. 5:97-100(1995).  
 RN [13]  
 RN VARIANT MPS-II VAL-346.  
 RX MEDLINE; 95322987.  
 RA LI P., HUFFMAN P., THOMPSON J.N.;  
 RA "Mutations of the iduronate-2-sulphatase gene on a T146T background in  
 RT three patients with Hunter syndrome.";  
 RL Hum. Mutat. 5:272-274(1995).  
 RN [14]  
 RN VARIANTS MPS-II.  
 RX MEDLINE; 97094177.  
 RA RATHMANN M., BUNGE S., BECK M., KRESSE H., TYLKI-SZYMANSKA A., GAL A.;  
 RA "Mucopolysaccharidosis type II (Hunter syndrome): mutation 'hot spots'  
 RT in the iduronate-2-sulphatase gene.";  
 RL Am. J. Hum. Genet. 59:1202-1209(1996).

RN [15]  
 RN VARIANTS MPS-II LEU-333 AND ASP-346.  
 RX MEDLINE; 96163494.  
 RA OLSEN T.C., EIKEN H.G., KNAPPSKOG P.M., KASE B.F., MANSSON J.-E.,  
 RA BOMAN H., APOLD J.;  
 RA "Mutations in the iduronate-2-sulphatase gene in five Norwegians with  
 RT Hunter syndrome.";  
 RL Hum. Genet. 97:198-203(1996).  
 RN [16]  
 RN VARIANTS MPS-II ASP-63; THR-347; GLN-468 AND LEU-468.  
 RX MEDLINE; 97365936.  
 RA VILLANI G.R.D., BALZANO N., GROSSO M., SALVADORE F., IZZO P.,  
 RA DI NATALE P.;  
 RA "Mucopolysaccharidosis type II: identification of six novel mutations  
 RT in Italian patients.";  
 RL Hum. Mutat. 10:71-75(1997).  
 RN [17]  
 RN VARIANT MPS-II GLN-468.  
 RX MEDLINE; 98041699.  
 RA SUKIGAWA K., SONG X.-Q., MASUNO M., FUKAO T., SHIMOZAWA N., FUKUDA S.,  
 RA ISOGAI K., NISHIO H., MATSUO M., TOMATSU S., KONDO N., ORII T.;  
 RA "Hunter disease in a girl caused by R468Q mutation in the  
 RT iduronate-2-sulphatase gene and skewed inactivation of the X  
 RT chromosome carrying the normal allele.";  
 RL Hum. Mutat. 10:361-367(1997).  
 RN [18]  
 RN VARIANTS MPS-II N-45; Y-115; L-228; R-266; K-434; K-485 AND C-502.  
 RX MEDLINE; 99092178.  
 RA VAFIADAKI E., COOPER A., HEPTINSTALL L.E., HATTON C.E., THORNLEY M.,  
 RA WRAITH J.E.;  
 RT "Mutation analysis in 57 unrelated patients with MPS II.";  
 RL Arch. Dis. Child. 79:237-241(1998).  
 RN [19]  
 RN VARIANTS MPS-II.  
 RX MEDLINE; 99118886.  
 RA KARSTEN S., VOSKOBOEVA E., TISHKANINA S., PETERSSON U.,  
 RA KRASNOPOLSKAJA X., BONDESON M.-L.;  
 RT "Mutational spectrum of the iduronate-2-sulphatase (IDS) gene in 36  
 RT unrelated Russian MPS II patients.";  
 RL Hum. Genet. 103:732-735(1998).  
 RN [20]  
 RN VARIANTS MPS-II LEU-86; HIS-88; PRO-88; ILE-118 AND HIS-266.  
 RA BALZANO N., VILLANI G.R.D., GROSSO M., IZZO P., DI NATALE P.;  
 RT "Detection of four novel mutations in the iduronate-2-sulphatase  
 RT gene.";  
 RL Hum. Mutat. 11:333-333(1998).  
 RN [21]  
 RN VARIANTS MPS-II THR-85; HIS-88; ILE-349 AND VAL-521.  
 RX MEDLINE; 98112423.  
 RA GORT L., COLL M.J., CHABAS A.;  
 RT "Mutations in the iduronate-2-sulphatase gene in 12 Spanish patients  
 RT with Hunter disease.";  
 RL Hum. Mutat. Suppl. 1:S66-S68(1998).  
 RN [22]  
 RN VARIANTS MPS-II PHE-143; TRP-184; VAL-269 AND HIS-348.  
 RA KARSTEN S.L., VOSKOBOEVA E., CARLBERG B.-M., KLEIJER W.J.,  
 RA TOENNESSEN T., PETERSSON U., BONDESON M.-L.;  
 RT "Identification of 9 novel gene mutations in 19 unrelated Hunter  
 RT syndrome (Mucopolysaccharidosis type II) patients.";  
 RL Hum. Mutat. 12:433-433(1998).  
 CC -!- FUNCTION: REQUIRED FOR THE LYSOSOMAL DEGRADATION OF HEPARAN  
 CC SULFATE AND DERMATAN SULFATE.  
 CC -!- CATALYTIC ACTIVITY: HYDROLYSIS OF THE 2-SULFATE GROUPS OF THE  
 CC L-IDURONATE 2-SULFATE UNITS OF DERMATAN SULFATE, HEPARAN SULFATE  
 CC AND HEPARIN.  
 CC -!- SUBUNIT: LIVER IDS IS COMPOSED OF TWO MAJOR FORMS (A AND B) WHICH  
 CC CONTAIN BOTH A 42 KD AND A 14 KD POLYPEPTIDES.  
 CC -!- SUBCELLULAR LOCATION: LYSOSOMAL.  
 CC -!- TISSUE SPECIFICITY: LIVER, KIDNEY, LUNG, AND PLACENTA.  
 CC -!- DISEASE: DEFECTS IN IDS ARE THE CAUSE OF HUNTER SYNDROME; AN X-  
 CC LINKED INBORN ERROR ALSO KNOWN AS MUCOPOLYSACCHARIDOSIS TYPE II  
 CC (MPS-II); LEADING TO LYSOSOMAL ACCUMULATION OF HEPARAN SULFATE AND  
 CC DERMATAN SULFATE AND THEIR EXCRETION IN URINE, MOST CHILDREN WITH

```

CC MPS-II HAVE A SEVERE FORM WITH EARLY SOMATIC ABNORMALITIES
CC INCLUDING SKELETAL DEFORMITIES, HEPATOSPLENOMEGALY, AND
CC PROGRESSIVE CARDIOPULMONARY DETERIORATION. A PROMINENT FEATURE IS
CC NEUROLOGICAL DAMAGE THAT PRESENTS AS DEVELOPMENTAL DELAY AND
CC HYPERACTIVITY BUT PROGRESSES TO MENTAL RETARDATION AND DEMENTIA.
CC THEY DIE BEFORE 15 YEARS OF AGE, USUALLY AS A RESULT OF
CC OBSTRUCTIVE AIRWAY DISEASE OR CARDIAC FAILURE. IN CONTRAST, THOSE
CC WITH A MILD FORM OF MPS-II MAY SURVIVE INTO ADULTHOOD, WITH
...: remainder of annotations omitted.

Query Match      100.0%; Score 4069; DB 1; Length 550;
Best Local Similarity 100.0%; Pred. No. 0.00e+00;
Matches 550; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Db 1 MPPRTGRLGLWGLVSSVCVAGLSETQANSTTDALNVLIIIVDDLRPSLCYGDKLYR 60
QY 1 MPPRTGRLGLWGLVSSVCVAGLSETQANSTTDALNVLIIIVDDLRPSLCYGDKLYR 60

Db 61 SPNIDQLASHSLLFQNAFAQAQVACAPSRVSLTGRRPDTRLYDFNSYWRVHAGNFSTIP 120
QY 61 SPNIDQLASHSLLFQNAFAQAQVACAPSRVSLTGRRPDTRLYDFNSYWRVHAGNFSTIP 120

Db 121 QYFKENGYYTMSVGVKPHFGISSNHTDSDPSYSPPPYHPSSEKYENTKTCRGPDELHA 180
QY 121 QYFKENGYYTMSVGVKPHFGISSNHTDSDPSYSPPPYHPSSEKYENTKTCRGPDELHA 180

Db 181 NLLCPVDVLDVPEGTLPDKQTEQAQLLEKMKTSASFFFLAVGYKHPHIPPYPKFOK 240
QY 181 NLLCPVDVLDVPEGTLPDKQTEQAQLLEKMKTSASFFFLAVGYKHPHIPPYPKFOK 240

Db 241 LYPLENITLAPDPEVDPGLPPVAYNPWMDIQREDVQALNISVPYGPVDFQKIRQSY 300
QY 241 LYPLENITLAPDPEVDPGLPPVAYNPWMDIQREDVQALNISVPYGPVDFQKIRQSY 300

Db 301 FASVSYLDQVGRLLSALDDQLANSTIIAFTSDHGALGEGEWAKYSNFDVATHVPLI 360
QY 301 FASVSYLDQVGRLLSALDDQLANSTIIAFTSDHGALGEGEWAKYSNFDVATHVPLI 360

Db 361 FYVPGRTASLPEAGEKLFYPLDPPDSASQLMPEGRQSDMLVELVSLFPTLAGLAGLVPP 420
QY 361 FYVPGRTASLPEAGEKLFYPLDPPDSASQLMPEGRQSDMLVELVSLFPTLAGLAGLVPP 420

Db 421 RCPVPSEHVELCRGKNNLKHFRDLIEDPYLPGNPRELIAYSQYPRSDIPQWNSDKP 480
QY 421 RCPVPSEHVELCRGKNNLKHFRDLIEDPYLPGNPRELIAYSQYPRSDIPQWNSDKP 480

Db 481 SLKDIKIMGYSIRTDYRYTWVGFNPDEFLANFSDIHAGELYFYVDSPLQDHNMYNDSQ 540
QY 481 SLKDIKIMGYSIRTDYRYTWVGFNPDEFLANFSDIHAGELYFYVDSPLQDHNMYNDSQ 540

Db 541 GGDLLQLIMP 550
QY 541 GGDLLQLIMP 550

RESULT 2
ID IDS_MOUSE STANDARD; PRT; 563 AA.
AC Q08890;
DT 01-OCT-1994 (Rel. 30, Created)
DT 01-OCT-1994 (Rel. 30, Last sequence update)
DT 01-OCT-1996 (Rel. 34, Last annotation update)
DE IDURONATE 2-SULFATASE PRECURSOR (EC 3.1.6.13).
GN IDS.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
OC Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=THYMUS;
RX MEDLINE; 93315172.
RA DANIELE A., FAUST C.J., HERMAN G.E., DI NATALE P., BALLABIO A.;
*Cloning and characterization of the cDNA for the murine iduronate

```

```

RT sulfatase gene.";
RL Genomics 16:755-757(1993).
CC FUNCTION: REQUIRED FOR THE LYSOSOMAL DEGRADATION OF HEPARAN
CC SULFATE AND DERMATAN SULFATE.
CC CATALYTIC ACTIVITY: HYDROLYSIS OF THE 2-SULFATE GROUPS OF THE
CC L-IDURONATE 2-SULFATE UNITS OF DERMATAN SULFATE, HEPARAN SULFATE
CC AND HEPARIN.
CC SUBCELLULAR LOCATION: LYSOSOMAL.
CC SIMILARITY: BELONGS TO THE SULFATASE FAMILY.
CC
CC THIS SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC
CC EMBL; L07921; AAA37880.1; -.
DR PIR; A47153; A47153.
DR MGD; MGI:96417; IDS.
DR PROSITE; PS00523; SULFATASE_1; 1.
DR PROSITE; PS00149; SULFATASE_2; 1.
DR PFAM; PF00884; Sulfatase; 1.
KW Hydrolase; Glycoprotein; Lysosome; zymogen; Signal.
FT SIGNAL 1 ? POTENTIAL.
FT PROPEP ? 59 BY SIMILARITY.
FT CHAIN 60 563 IDURONATE 2-SULFATASE.
FT MOD_RES 99 99 2-AMINO-3-OXOPROPIONIC ACID (BY
FT SIMILARITY).
FT CARBOHYD 130 130 POTENTIAL.
FT CARBOHYD 159 159 POTENTIAL.
FT CARBOHYD 261 261 POTENTIAL.
FT CARBOHYD 295 295 POTENTIAL.
FT CARBOHYD 526 526 POTENTIAL.
FT CARBOHYD 550 550 POTENTIAL.
SQ SEQUENCE 563 AA; 63437 MW; 9A6F7CD9 CRC32;

Query Match      76.9%; Score 3129; DB 1; Length 563;
Best Local Similarity 79.0%; Pred. No. 0.00e+00;
Matches 433; Conservative 58; Mismatches 53; Indels 4; Gaps 4;

Db 19 ARAIWQLSFLSLGSECTALESAAAGNSATDAILNILLIIVDDLRPSLCYGDKLVSPN 78
QY 4 PRTGRLLGLVSSVCVAGLSETQANSTTDALNVLIIIVDDLRPSLCYGDKLVSPN 63

Db 79 IDOLASHSLVLFQNAFAQAQVACAPSRVSLTGRRPDTRLYDFNSYWRVHAGNFSTIPQYF 138
QY 64 IDOLASHSLVLFQNAFAQAQVACAPSRVSLTGRRPDTRLYDFNSYWRVHAGNFSTIPQYF 123

Db 139 KENGYVTMSVGVKPHFGISSNHSDDYPYSPSFPYHPSSEKYENTKTCQDGKLANLL 198
QY 124 KENGYVTMSVGVKPHFGISSNHTDSDPSYSPSFPYHPSSEKYENTKTCRPGDELHANLL 183

Db 199 CPVDVADVPEGTLPDKQSTEEAIRLLEKMKTSGPSFFLAVGYKHPHIPPYPKFOKLYP 258
QY 184 CPVDVADVPEGTLPDKQSTEEAIRLLEKMKTSASPPFLAVGYKHPHIPPYPKFOKLYP 243

Db 259 LENITLAPDPEVDPGLPPVAYNPWMDIQREDVQALNISVPYGPVDFQKIRQSYFAS 318
QY 244 LENITLAPDPEVDPGLPPVAYNPWMDIQREDVQALNISVPYGPVDFQKIRQSYFAS 303

Db 319 VSYLDQVGHVLSALDDQLANSTIIAFTSDHGALGEGEWAKYSNFDVATHVPLMLYV 378
QY 304 VSYLDQVGHVLSALDDQLANSTIIAFTSDHGALGEGEWAKYSNFDVATHVPLIYV 363

Db 379 PGRTAPLPAAGKLFYPLDPPDSASQLMPEGRQSDMLVELVSLFPTLAGLAGLVLAGP 438
QY 364 PGRTASLPEAGEKLFYPLDPPDSASQLMPEGRQSDMLVELVSLFPTLAGLAGLVPRCP 423

Db 439 SLFLMPLSFAEKAR-IFRSTIC-SSMTWKRSTCLVPRELIAYSQYPRPADPQWNSDKPTL 496
QY 424 VPSFHVLCREGKNNLKHFRDLIEDPYLPGNPRELIAYSQYPRSDIPQWNSDKPSL 482

```



QY 61 SPNIDQLASHLLFQNAFAQAQVACPSRVSLTGR-RPDTTRL-YD-FNSY-WRVHAGNF 116

Db 120 PPSEVTMARLLKRGYATALLICK-WHLGLSCRGATDFCHHPLRHGDFRPLGVPTT--NLR 176

QY 117 --S--TIQYKENGCVYTVMSGVKVPHPGISS-NHTD--DSPYSWSPFPVHPSEKVENTK 169

Db 177 DCRGAGTVFGALRVFAAGPLAALGASLAAMAARWAGLARPVWALAGTAGTAAMLA VGG 236

QY 170 TCR-GPDGELHANL-LCPVDVLDVPEGLTPDKOSTE-QAI-QLLEKMKTSASPFPLAVGY 225

Db 237 PRSASCLGF-RP-ANCFMLDDLAQRDTYGLLFRRLADEAALFLRRNRPFLFLSLF 294

QY 226 HKPHIPFPYKFEQKLYPLENITLAPDEVPDGLPP-VAYNPWMDIRQ-RE-D-VQALN- 280

Db 295 LHVHTAHADPGFAGRSLSHGAGDSVEEMDGVGVLAALDELGLARETLVYVFTSDHGAH 354

QY 281 ISVPYGPPI-VDFQ-RKIQSFASVSLDTQVGRLLSALDQLANSTIAFTSDHGW 338

Db 355 VEELGPRGE--RMG-GSN-G-VF-RGGKGNW-EGGVRY-PCLVRW--PREL-SGRVVA 403

QY 339 LGEHGEWAKSYNFVATHVPLIFYVPGRTASLPEAGEKLPYLDPDFDSASQLMEPGQSM 398

Db 404 EPTSLMDVFPVARLAGAELP 424

QY 399 DLVELVSLFPTLAGLAGLQVP 419

RESULT 5

ID ARSF\_HUMAN STANDARD; PRT; 591 AA.

AC P54793;

DT 01-OCT-1996 (Rel. 34, Created)

DT 15-JUL-1999 (Rel. 38, Last sequence update)

DT 15-JUL-1999 (Rel. 38, Last annotation update)

DE ARYLSULFATASE F PRECURSOR (EC 3.1.6.-) (ASF).

GN ARSF.

OS Homo sapiens (Human).

OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;

OC Eutheria; Primates; Catarrhini; Homnidae; Homo.

RN [1]

RP SEQUENCE FROM N.A.

RC TISSUE-KIDNEY;

RX MEDLINE; 97336043.

RA PUC A.A., ZOLLO M., REPETTO M., ANDOLFI G., GUFFANTI A., SIMON G.,

RA BALLABIO A., FRANCO B.;

RT "Identification by shotgun sequencing, genomic organization, and

RT functional analysis of a fourth arylsulfatase gene (ARSF) from the

RT xp22.3 region."

RL Genomics 42:192-199(1997).

RN [2]

RP SEQUENCE OF 370-423 FROM N.A.

RC TISSUE-KIDNEY;

RX MEDLINE; 95236447.

RA FRANCO B., MERONI G., PARENTI G., LEVILLIERS J., BERNARD L.,

RA GEBBIA M., COX L., MAROTEAUX P., SHEFFIELD L., RAPPOLD G.A.,

RA ANDRIA G., PETIT C., BALLABIO A.;

RT "A cluster of sulfate genes on xp22.3: mutations in

RT chondrodysplasia punctata (CDPX) and implications for warfarin

RT embryopathy."

RL Cell 81:15-25(1995).

CC -1- ENZYME REGULATION: NOT INHIBITED BY DHEAS OR WARFARIN.

CC -1- MISCELLANEOUS: OPTIMUM PH IS 8.

CC -1- SIMILARITY: BELONGS TO THE SULFATASE FAMILY.

CC -----

CC This SWISS-PROT entry is copyright. It is produced through a collaboration

CC between the Swiss Institute of Bioinformatics and the EMBL outstation -

CC the European Bioinformatics Institute. There are no restrictions on its

CC use by non-profit institutions as long as its content is in no way

CC modified and this statement is not removed. Usage by and for commercial

CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>

CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).

CC -----

DR EMBL; X97868; CAA66462.1; -

DR HSSP; P15289; LAUK.

MM; 300003; -

DR PROSITE; PS00523; SULFATASE\_1; 1.

DR PROSITE; PS00149; SULFATASE\_2; 1.

KW Hydrolase; Glycoprotein; Signal.

FT SIGNAL 1 19 POTENTIAL.

FT CHAIN 20 591 ARYLSULFATASE F.

FT MOD\_RES 79 79 2-AMINO-3-OXOPROPIONIC ACID (BY

FT SIMILARITY).

FT ACT\_SITE 140 140 BY SIMILARITY.

FT CARBOHYD 51 51 POTENTIAL.

FT CARBOHYD 118 118 POTENTIAL.

FT CARBOHYD 337 337 POTENTIAL.

SQ SEQUENCE 591 AA; 66003 MM; 233649DS CRC32;

Query Match 6.38; Score 258; DB 1; Length 591;

Best Local Similarity 27.98; Pred. No. 6.75e-29;

Matches 97; Conservative 86; Mismatches 133; Indels 32; Gaps 24;

Db 7 LVFMSLVALLNTPWGHTGCMTRPNIVLWDDLGIGDLCYGNDRTPHIDRLAREG 66

QY 15 LV-LSSVCVALGS-ETQANSTTDALNVLIIIVDDL-RPSLCYGDKLVSPNIDQLASHS 71

Db 67 VRLTQHISRASLCSPRSFAELTGRTPIRSQWVS-SGNRRVQNLAVPAGLPLNETTLAAL 125

QY 72 LLFQNAFAQAQVACPSRVSLTGRPDTRILYDFNSYRVV-H-----AG-NF--STIPQY 122

Db 126 LKQGYSTGLIGK-WHQGLNCDRSDOCHHPYNGFYDYGMPFTLVDSQWPDSPRNTL 184

QY 123 FKENGVTYSGVKVFPHPGISSN-HTDSD--PYSWSFPYHPSPSEKY-EN--TKTCRGPDG 176

Db 185 AFESQWLCL-VOLVAITAILTLTKGLSGWVSPVLLIFSLIFLGLAGYAWFSHTSPLY 243

QY 177 ELHANL-LCPVDVLDVPEGLTP-DKOSTEQAIQLLEKMKTSASPFPLAVGYHKPHIDRY 234

Db 244 -WDCL-LMRGHEITEQPMKAERAGSIMVKEAISPLE-RHSEKETFLFFSLHVHTPLPTT 300

QY 235 PKFEQKLYPLENITLAP-DEVPDGLPPVAYNPWMDIRQEDVQALNIS-VP-YGPIPV- 290

Db 301 DDFGTGSKHGLYGDNVEEMDSWVGKILDAIDDFGLRNNTLYVFTSDHG 348

QY 291 -DFQRKIRQS-YEASVSYLDQVGRLLSALDQLANSTIAFTSDHG 336

RESULT 6

ID ARSF\_HUMAN STANDARD; PRT; 589 AA.

AC P51690;

DT 01-OCT-1996 (Rel. 34, Created)

DT 01-OCT-1996 (Rel. 34, Last sequence update)

DT 15-JUL-1999 (Rel. 38, Last annotation update)

DE ARYLSULFATASE E PRECURSOR (EC 3.1.6.-) (ASE).

GN ARSF.

OS Homo sapiens (Human).

OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;

OC Eutheria; Primates; Catarrhini; Homnidae; Homo.

RN [1]

RP SEQUENCE FROM N.A., AND VARIANTS CDPX.

RC TISSUE-KIDNEY;

RX MEDLINE; 95236447.

RA FRANCO B., MERONI G., PARENTI G., LEVILLIERS J., BERNARD L.,

RA GEBBIA M., COX L., MAROTEAUX P., SHEFFIELD L., RAPPOLD G.A.,

RA ANDRIA G., PETIT C., BALLABIO A.;

RT "A cluster of sulfate genes on xp22.3: mutations in

RT chondrodysplasia punctata (CDPX) and implications for warfarin

RT embryopathy."

RL Cell 81:15-25(1995).

RN [2]

RP VARIANT CDPX TVR-492.

RX MEDLINE; 98072521.

RA PARENTI G., BUTTINI P., MERONI G., FRANCO B., BERNARD L.,

RA RIZZOLO M.G., BRUNETTI-PIERRI N., BALLABIO A., ANDRIA G.;

RT "X-linked recessive chondrodysplasia punctata due to a new point

RT mutation of the ARSE gene."

RL Am. J. Med. Genet. 73:139-143(1997).





AC P34059;  
DT 01-FEB-1994 (Rel. 28, Created)  
DT 01-FEB-1994 (Rel. 28, Last sequence update)  
DT 15-JUL-1998 (Rel. 36, Last annotation update)  
DE N-ACETYLALACTOSAMINE-6-SULFATASE PRECURSOR (EC 3.1.6.4) (N-  
DE ACETYLALACTOSAMINE-6-SULFATE SULFATASE) (GALACTOSE-6-SULFATE  
DE SULFATASE) (GALNAC6S SULFATASE) (CHONDROITINSULFATASE)  
DE (CHONDROITINASE).  
GN GALNS.  
OS Homo sapiens (Human).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;  
OC Eutheria; Primates; Catarrhini; Homnidae; Homo.  
RN [1]  
RN SEQUENCE FROM N.A., AND PARTIAL SEQUENCE.  
RC TISSUE-PLACENTA;  
RX MEDLINE; 92095973.  
RA TOMATSU S., FUKUDA S., MASUE M., SUKEGAWA K., FUKAO T., YAMAGISHI A.,  
RA HORI T., IWATA H., OGAWA T., NAKASHIMA Y., HANYU Y., HASHIMOTO T.,  
RA TITANI K., OYAMA R., SUZUKI M., YAGI K., HAYASHI Y., ORII T.;  
RT "Morquio disease: isolation, characterization and expression of full-  
RT length cDNA for human N-acetylglactosamine-6-sulfate sulfatase.";  
RT Biochem. Biophys. Res. Commun. 181:677-683(1991).  
RN [2]  
RN SEQUENCE FROM N.A.  
RX MEDLINE; 95095267.  
RA MORRIS C.P., GUO X.H., APOSTOLOU S., HOPWOOD J.J., SCOTT H.S.;  
RT "Morquio A syndrome: cloning, sequence, and structure of the human N-  
RT acetylglactosamine 6-sulfatase (GALNS) gene.";  
RL Genomics 22:652-654(1994).  
RN [3]  
RN VARIANT MPS-IVA LYS-204.  
RX MEDLINE; 92395122.  
RA FUKUDA S., TOMATSU S., MASUE M., SUKEGAWA K., IWATA H., OGAWA T.,  
RA NAKASHIMA Y., HORI T., YAMAGISHI A., HANYU Y., MOROOKA K., KIMAN T.,  
RA HASHIMOTO T., ORII T.;  
RT "Mucopolysaccharidosis type IVA. N-acetylglactosamine-6-sulfate  
RT sulfatase exonic point mutations in classical Morquio and mild  
RT cases.";  
RL J. Clin. Invest. 90:1049-1053(1992).  
RN [4]  
RN VARIANTS MPS-IVA.  
RX MEDLINE; 95397840.  
RA TOMATSU S., FUKUDA S., COOPER A., WRAITH J.E., MARUF REZVI G.,  
RA YAMAGISHI A., YAMADA N., KATO Z., ISOGAI K., SUKEGAWA K., KONDO N.,  
RA SUZUKI Y., SHIMOZAWA N., ORII T.;  
RT "Mucopolysaccharidosis IVA: identification of a common missense  
RT mutation I113F in the N-Acetylglactosamine-6-sulfate sulfatase  
RT gene.";  
RL Am. J. Hum. Genet. 57:556-563(1995).  
RN [5]  
RN VARIANTS MPS-IVA.  
RX MEDLINE; 95315929.  
RA OGAWA T., TOMATSU S., FUKUDA S., YAMAGISHI A., MARUF REZVI G.,  
RA SUKEGAWA K., KONDO N., SUZUKI Y., SHIMOZAWA N., ORII T.;  
RT "Mucopolysaccharidosis IVA: screening and identification of mutations  
RT of the N-acetylglactosamine-6-sulfate sulfatase gene.";  
RL Hum. Mol. Genet. 4:341-349(1995).  
RN [6]  
RN VARIANTS MPS-IVA VAL-77; TRP-90; VAL-96; LEU-151; GLY-230 & THR-291.  
RX MEDLINE; 95359983.  
RA TOMATSU S., FUKUDA S., COOPER A., WRAITH J.E., MARUF REZVI G.,  
RA YAMAGISHI A., YAMADA N., KATO Z., ISOGAI K., SUKEGAWA K., KONDO N.,  
RA SUZUKI Y., SHIMOZAWA N., ORII T.;  
RT "Mucopolysaccharidosis type IVA: identification of six novel  
RT mutations among non-Japanese patients.";  
RL Hum. Mol. Genet. 4:741-743(1995).  
RN [7]  
RN VARIANT MPS-IVA SER-487.  
RX MEDLINE; 96047158.  
RA TOMATSU S., FUKUDA S., COOPER A., WRAITH J.E., YAMADA N., ISOGAI K.,  
RA KATO Z., SUKEGAWA K., KONDO N., SUZUKI Y., SHIMOZAWA N., ORII T.;  
RT "Two new mutations, Q473X and N487S, in a Caucasian patient with  
RT mucopolysaccharidosis IVA (Morquio disease).";  
RL Hum. Mutat. 6:195-196(1995).  
RN [8]  
RN VARIANTS MPS-IVA ALA-138; SER-151 AND LEU-151.  
RX MEDLINE; 96216146.  
RA TOMATSU S., FUKUDA S., YAMAGISHI A., COOPER A., WRAITH J.E., HORI T.,  
RA KATO Z., YAMADA N., ISOGAI K., SUKEGAWA K., KONDO N., SUZUKI Y.,  
RA SHIMOZAWA N., ORII T.;  
RT "Mucopolysaccharidosis IVA: four new exonic mutations in patients  
RT with N-acetylglactosamine-6-sulfate sulfatase deficiency.";  
RL Am. J. Hum. Genet. 58:950-962(1996).  
RN [9]  
RN VARIANTS MPS-IVA CYS-94 AND VAL-97.  
RX MEDLINE; 96423834.  
RA COLE D.E.C., FUKUDA S., GORDON B.A., RIP J.W., LECOUTEUR A.N.,  
RA RUPAR C.A., TOMATSU S., OGAWA T., SUKEGAWA K., ORII T.;  
RT "Heteroallelic missense mutations of the galactosamine-6-sulfate  
RT sulfatase (GALNS) gene in a mild form of Morquio disease (MPS IVA).";  
RL Am. J. Med. Genet. 63:558-565(1996).  
RN [10]  
RN VARIANTS MPS-IVA.  
RX MEDLINE; 97442274.  
RA BUNGE S., KLEIJER W.J., TYLKI-SZYMANSKA A., STEGLICH C., BECK M.,  
RA TOMATSU S., FUKUDA S., POORTHUIS B.J.H.M., CZARTORYSKA B., ORII T.,  
RA GAL A.;  
RT "Identification of 31 novel mutations in the N-acetylglactosamine-6-  
RT sulfatase gene reveals excessive allelic heterogeneity among patients  
RT with Morquio A syndrome.";  
RL Hum. Mutat. 10:223-232(1997).  
RN [11]  
RN VARIANTS MPS-IVA.  
RX MEDLINE; 98041700.  
RA TOMATSU S., FUKUDA S., COOPER A., WRAITH J.E., FERREIRA P.,  
RA DI NATALE P., TORTORA P., FUJIMOTO A., KATO Z., YAMADA N., ISOGAI K.,  
RA YAMAGISHI A., SUKEGAWA K., SUZUKI Y., SHIMOZAWA N., KONDO N.,  
RA SLY W.S., ORII T.;  
RT "Fourteen novel mucopolysaccharidosis IVA producing mutations in  
RT GALNS gene.";  
RL Hum. Mutat. 10:368-375(1997).  
RN [12]  
RN VARIANTS MPS-IVA.  
RX MEDLINE; 98180718.  
RA YAMADA N., FUKUDA S., TOMATSU S., MULLER V., HOPWOOD J.J., NELSON J.,  
RA KATO Z., YAMAGISHI A., SUKEGAWA K., KONDO N., ORII T.;  
RT "Molecular heterogeneity in mucopolysaccharidosis IVA in Australia  
RT and Northern Ireland: nine novel mutations including T312S, a common  
RT allele that confers a mild phenotype.";  
RL Hum. Mutat. 11:202-208(1998).  
RN [13]  
RN VARIANTS SER-393 AND MET-488.  
RX MEDLINE; 98112415.  
RA TOMATSU S., FUKUDA S., COOPER A., WRAITH J.E., YAMAGISHI A., KATO Z.,  
RA YAMADA N., ISOGAI K., SUKEGAWA K., SUZUKI Y., SHIMOZAWA N., KONDO N.,  
RA ORII T.;  
RT "Fifteen polymorphisms in the N-acetylglactosamine-6-sulfate  
RT sulfatase (GALNS) gene: diagnostic implications in Morquio disease.";  
RL Hum. Mutat. Suppl. 1:S42-S46(1998).  
CC -!- CATALYTIC ACTIVITY: HYDROLYSIS OF THE 6-SULFATE GROUPS OF THE  
CC N-ACETYL-D-GALACTOSAMINE 6-SULFATE UNITS OF CHONDROITIN SULFATE  
CC AND OF THE D-GALACTOSE 6-SULFATE UNITS OF KERATAN SULFATE.  
CC -!- SUBUNIT: OLIGOMER OF DISULFIDE LINKED 40- AND 15-KD POLYPEPTIDES.  
CC -!- SUBCELLULAR LOCATION: LYSOSOMAL.  
CC -!- DISEASE: DEFECTS IN GALNS ARE THE CAUSE OF MUCOPOLYSACCHARIDOSIS  
CC TYPE IVA (MPS-IVA) (ALSO KNOWN AS MORQUIO A SYNDROME) WHICH IS  
CC CHARACTERIZED BY SPECIFIC SPONDYLOEPHYSEAL DYSPLASIA, SHORT  
CC TRUNK DWARFISM, COXA VALGA, ODONTOID HYPOPLASIA, CORNEAL  
CC OPACITIES, PRESERVATION OF INTELLIGENCE, AND EXCESSIVE URINARY  
CC EXCRETION OF KERATAN SULFATE AND CHONDROITIN-6-SULFATE. SEVERELY  
CC AFFECTED PATIENTS USUALLY DIE OF CARDIOPULMONARY DISTURBANCE OR  
CC CERVICAL CORD COMPRESSION IN THE SECOND OR THIRD DECADE OF LIFE.  
CC -!- SIMILARITY: BELONGS TO THE SULFATASE FAMILY.  
CC -----  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -







DT 15-JUL-1999 (Rel. 38, Last annotation update)  
DE ARYLSULFATASE A PRECURSOR (EC 3.1.6.8) (ASA) (CEREBROSIDE-SULFATASE).  
GN ARSA.  
OS Homo sapiens (Human).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;  
OC Eutheria; Primates; Catarrhini; Homnidae; Homo.  
RN [1]  
RN SEQUENCE FROM N.A.  
RX MEDLINE; 90361046.  
RA KREYSING J., VON FIGURA K., GIESELMANN V.;  
RT "Structure of the arylsulfatase A gene.";  
RL Eur. J. Biochem. 191:627-631(1990).  
RN [2]  
RN SEQUENCE FROM N.A.  
RX MEDLINE; 89093115.  
RA STEIN C., GIESELMANN V., KREYSING J., SCHMIDT B., POHLMANN R.,  
RA WAHEED A., MEYER H.E., O'BRIEN J.S., VON FIGURA K.;  
RT "Cloning and expression of human arylsulfatase A";  
RL J. Biol. Chem. 264:1252-1259(1989).  
RN [3]  
RN SEQUENCE FROM N.A.  
RX ADAMS M.D., KERLAVAGE A.R., FULDNER R.A., PHILLIPS C.A., VENTER J.C.;  
RL Submitted (JUL-1996) to the EMBL/GenBank/DBJ databases.  
RN [4]  
RN SEQUENCE OF 19-33 AND 434-479, AND SUBUNITS.  
RX MEDLINE; 92338230.  
RA FUJII T., KOBAYASHI T., HONKE K., GASA S., ISHIKAWA M., SHIMITSU T.,  
RA MAKITA A.;  
RT "Proteolytic processing of human lysosomal arylsulfatase A";  
RL Biochim. Biophys. Acta 1122:93-98(1992).  
RN [5]  
RN PARTIAL SEQUENCE, AND 2-AMINO-3-OXOPROPIONIC ACID MODIFICATION SITE.  
RX MEDLINE; 95354208.  
RA SCHMIDT B., SELMER T., INGENDOH A., VON FIGURA K.;  
RT "A novel amino acid modification in sulfatases that is defective in  
RT multiple sulfatase deficiency.";  
RL Cell 82:271-278(1995).  
RN [6]  
RN X-RAY CRYSTALLOGRAPHY (2.1 ANGSTROMS).  
RX MEDLINE; 98191325.  
RA LUKATELA G., KRAUSS N., THEIS K., SELMER T., GIESELMANN V.,  
RA VON FIGURA K., SAENGER W.;  
RT "Crystal structure of human arylsulfatase A: the aldehyde function  
RT and the metal ion at the active site suggest a novel mechanism for  
RT sulfate ester hydrolysis.";  
RL Biochemistry 37:3654-3664(1998).  
RN [7]  
RN REVIEW ON MLD VARIANTS.  
RX MEDLINE; 95170731.  
RA GIESELMANN V., ZLOTOGORA J., HARRIS A., WENGER D.A., MORRIS C.P.;  
RT "Molecular genetics of metachromatic leukodystrophy.";  
RL Hum. Mutat. 4:233-242(1994).  
RN [8]  
RN VARIANT MLD GLN-84.  
RX MEDLINE; 92344341.  
RA KAPPLER J., VON FIGURA K., GIESELMANN V.;  
RT "Late-onset metachromatic leukodystrophy: molecular pathology in two  
RT siblings.";  
RL Ann. Neurol. 31:256-261(1992).  
RN [9]  
RN VARIANT MLD PHE-96.  
RX MEDLINE; 91328147.  
RA GIESELMANN V., FLUHARTY A.L., TONNESSEN T., VON FIGURA K.;  
RT "Mutations in the arylsulfatase A pseudodeficiency allele causing  
RT metachromatic leukodystrophy.";  
RL Am. J. Hum. Genet. 49:407-413(1991).  
RN [10]  
RN VARIANT MLD ASP-99.  
RX MEDLINE; 91206410.  
RA KONDO R., WAKAMATSU N., YOSHINO H., FUKUHARA N., MIYATAKE T.,  
RA TSUJII S.;  
RT "Identification of a mutation in the arylsulfatase A gene of a  
RT patient with adult-type metachromatic leukodystrophy.";

RL Am. J. Hum. Genet. 48:971-978(1991).  
RN [11]  
RN VARIANT MLD LEU-426, AND VARIANTS CYS-193 AND SER-391.  
RX MEDLINE; 91074201.  
RA POLTEN A., FLUHARTY A.L., FLUHARTY C.B., KAPPLER J., VON FIGURA K.,  
RA GIESELMANN V.;  
RT "Molecular basis of different forms of metachromatic leukodystrophy.";  
RL New Engl. J. Med. 324:18-22(1991).  
RN [12]  
RN VARIANT MLD SER-122.  
RX MEDLINE; 94063853.  
RA HONKE K., KOBAYASHI T., FUJII T., GASA S., XU M., TAKAMARU Y.,  
RA KONDO R., TSUJII S., MAKITA A.;  
RT "An adult-type metachromatic leukodystrophy caused by substitution of  
RT serine for glycine-122 in arylsulfatase A.";  
RL Hum. Genet. 92:451-456(1993).  
RN [13]  
RN VARIANT MLD LEU-136.  
RX MEDLINE; 95163939.  
RA KAFERT S., HEINISCH U., ZLOTOGORA J., GIESELMANN V.;  
RT "A missense mutation P136L in the arylsulfatase A gene causes  
RT instability and loss of activity of the mutant enzyme.";  
RL Hum. Genet. 95:201-204(1995).  
RN [14]  
RN VARIANT MLD ARG-245.  
RX MEDLINE; 93319632.  
RA HASEGAWA Y., KAWANE H., ETO Y.;  
RT "Mutations in the arylsulfatase A gene of Japanese patients with  
RT metachromatic leukodystrophy.";  
RL DNA Cell Biol. 12:493-498(1993).  
RN [15]  
RN VARIANT MLD MET-274.  
RX MEDLINE; 94004907.  
RA HARVEY J.S., NELSON P.V., CAREY W.F., ROBERTSON E.F., MORRIS C.P.;  
RT "An arylsulfatase A (ARSA) missense mutation (T274M) causing late-  
RT infantile metachromatic leukodystrophy.";  
RL Hum. Mutat. 2:261-267(1993).  
RN [16]  
RN VARIANT MLD SER-309.  
RX MEDLINE; 93318834.  
RA KREYSING J., BOHNE W., BOSENBERG C., MARCHESINI S., TURPIN J.C.,  
RA BAUMANN N., VON FIGURA K., GIESELMANN V.;  
RT "High residual arylsulfatase A (ARSA) activity in a patient with  
RT late-infantile metachromatic leukodystrophy.";  
RL Am. J. Hum. Genet. 53:339-346(1993).  
RN [17]  
RN VARIANT SER-350.  
RX MEDLINE; 90083282.  
RA GIESELMANN V., POLTEN A., KREYSING J., VON FIGURA K.;  
RT "Arylsulfatase A pseudodeficiency: loss of a polyadenylation signal  
RT and N-glycosylation site.";  
RL Proc. Natl. Acad. Sci. U.S.A. 86:9436-9440(1989).  
RN [18]  
RN VARIANT MLD LEU-426.  
RX MEDLINE; 93202658.  
RA BARTH M.L., FENSOM A., HARRIS A.;  
RT "Prevalence of common mutations in the arylsulfatase A gene in  
RT metachromatic leukodystrophy patients diagnosed in Britain.";  
RL Hum. Genet. 91:73-77(1993).  
RN [19]  
RN VARIANTS MLD VAL-212; VAL-224 AND TYR-295.  
RX MEDLINE; 94154687.  
RA BARTH M.L., FENSOM A., HARRIS A.;  
RT "Missense mutations in the arylsulfatase A genes of metachromatic  
RT leukodystrophy patients.";  
RL Hum. Mol. Genet. 2:2117-2121(1993).  
RN [20]  
RN VARIANTS MLD LEU-82; TYR-172; CYS-201; GLN-311; VAL-335 AND TRP-390.  
RX MEDLINE; 96047150.  
RA BARTH M.L., FENSOM A., HARRIS A.;  
RT "Identification of seven novel mutations associated with  
RT metachromatic leukodystrophy.";  
RL Hum. Mutat. 6:170-176(1995).



[10]  
RN VARIANTS MPS-VI TRP-152 AND GLN-160.  
RX MEDLINE: 94171224.  
RA VOSKOBOEVA E., ISBRANDT D., VON FIGURA K., KRASNOPOLSKAYA X.,  
RA PETERS C.;  
RT "Four novel mutant alleles of the arylsulfatase B gene in two  
RT patients with intermediate form of mucopolysaccharidosis VI  
RT (Maroteaux-Lamy syndrome).";  
RL Hum. Genet. 93:259-264 (1994).  
RN [11]  
RP VARIANT MPS-IV ARG-302.  
RA VILLANI G.R.D., BALZANO N., DI NATALE P.;  
RT "Two novel mutations of the arylsulfatase B gene in two Italian  
RT patients with severe form of mucopolysaccharidosis.";  
RL Hum. Mutat. 11:410-410(1998).  
CC -1- CATALYTIC ACTIVITY: HYDROLYSIS OF THE 4-SULFATE GROUPS OF THE  
CC N-ACETYL-D-GALACTOSAMINE 4-SULFATE UNITS OF CHONDROITIN SULFATE  
CC AND DERMATAN SULFATE.  
CC -1- SUBUNIT: MONOMER.  
CC -1- SUBCELLULAR LOCATION: LYSOSOMAL.  
CC DISEASE: DEFECTS IN ARSB ARE THE CAUSE OF MAROTEAUX-LAMY  
CC SYNDROME; ALSO KNOWN AS MUCOPOLYSACCHARIDOSIS TYPE VI (MPS-VI).  
CC THIS DISEASE IS CHARACTERIZED BY THE ACCUMULATION OF DERMATAN  
CC SULFATE IN LYSOSOMES. CLINICAL FEATURES CAN INCLUDE ABNORMAL  
CC GROWTH, SHORT STATURE, STIFF JOINTS, SKELETAL MALFORMATIONS,  
CC CORNEAL CLOUDING, HEPATOSPLENOMEGALY, AND CARDIAC ABNORMALITIES.  
CC A WIDE VARIATION IN CLINICAL SEVERITY IS OBSERVED.  
CC -1- DISEASE: MULTIPLE SULFATASE DEFICIENCY (MSD) IS A DISORDER THAT  
CC COMBINES FEATURES OF METACHROMATIC LEUKODYSTROPHY AND OF  
CC MUCOPOLYSACCHARIDOSIS; IT IS CHARACTERIZED BY A DECREASED  
CC ACTIVITY OF ALL KNOWN SULFATASES. IT SEEMS TO BE CAUSED FROM THE  
CC LACK OF POST-TRANSLATIONAL MODIFICATION OF A CYSTEINE INTO 2-  
CC AMINO-3-OXOPROPIONIC ACID.  
CC -1- SIMILARITY: BELONGS TO THE SULFATASE FAMILY.  
CC -----  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC -----  
DR EMBL; J05225; AAA51784.1; -;  
DR EMBL; M32373; AAA51779.1; -;  
DR EMBL; X72735; CAA51272.1; -;  
DR EMBL; X72736; CAA51272.1; JOINED.  
DR EMBL; X72737; CAA51272.1; JOINED.  
DR EMBL; X72738; CAA51272.1; JOINED.  
DR EMBL; X72739; CAA51272.1; JOINED.  
DR EMBL; X72740; CAA51272.1; JOINED.  
DR EMBL; X72741; CAA51272.1; JOINED.  
DR EMBL; X72742; CAA51272.1; JOINED.  
DR EMBL; S57777; AAB19988.1; -;  
DR PIR; A35078; A35078.  
DR PDB; 1FSU; 04-FEB-98.  
DR MIM; 253200; -;  
DR MIM; 272200; -;  
DR PROSITE; PS00523; SULFATASE\_1; 1.  
DR PROSITE; PS00149; SULFATASE\_2; 1.  
DR PIR; PF00884; Sulfatase; 1.  
KW Hydrolase; Signal; Glycoprotein; Lysosome; Mucopolysaccharidosis;  
KW Disease mutation; Polymorphism; 3D-structure.  
FT SIGNAL 1 36 (POTENTIAL).  
FT CHAIN 37 533 ARYL-SULFATASE B.  
FT ACT\_SITE 147 147 POTENTIAL.  
FT MOD\_RES 91 91 2-AMINO-3-OXOPROPIONIC ACID.  
FT DISULFID 117 521  
FT DISULFID 121 155  
FT DISULFID 181 192  
FT DISULFID 405 447  
FT CARBOHYD 188 188  
FT CARBOHYD 279 279  
POTENTIAL.

FT CARBOHYD 291 291 PROBABLE.  
FT CARBOHYD 366 366  
FT CARBOHYD 426 426  
FT CARBOHYD 458 458  
FT VARIANT 92 92 POTENTIAL.  
FT VARIANT 95 95 T -> M (IN MPS-VI; MILD FORM).  
FT VARIANT 117 117 R -> Q (IN MPS-VI; MILD/SEVERE FORM).  
FT VARIANT 137 137 /FTid=VAR\_007294.  
FT VARIANT 152 152 /FTid=VAR\_007295.  
FT VARIANT 160 160 C -> R (IN MPS-VI; SEVERE FORM).  
FT VARIANT 210 210 /FTid=VAR\_007296.  
FT VARIANT 236 236 G -> V (IN MPS-VI; INTERMEDIATE FORM).  
FT VARIANT 302 302 /FTid=VAR\_007297.  
FT VARIANT 376 376 R -> W (IN MPS-VI; INTERMEDIATE FORM).  
FT VARIANT 393 393 /FTid=VAR\_007298.  
FT VARIANT 405 405 R -> Q (IN MPS-VI; INTERMEDIATE FORM).  
FT VARIANT 498 498 /FTid=VAR\_007299.  
FT CONFLICT 358 358 Y -> C (IN MPS-VI; MILD/INTERMEDIATE).  
FT SEQUENCE 533 AA; 59687 MW; 986B4C42 CRC32; V -> M (IN REF. 3).  
Query Match 4.7%; Score 190; DB 1; Length 533;  
Best Local Similarity 27.1%; Pred. No. 2.04e-15;  
Matches 69; Conservative 70; Mismatches 95; Indels 21; Gaps 17;  
Db 9 LPRGPGRRLL-LPVVPLLLLLLLAPPGSGAGARRPHLVFLVLLADLGNWVGFGHSR- 66  
QY 1 MPPRTGRGLLGLVLSVVAL-GSETQANSTDALNVLIIIVDDL-RPSLGCYGDKL 58  
Db 67 IRTPHDLAAGVLLDNYTQ-PLCTPSRSLTGTGYQRTGL-QHIIWPCQSPCVPL 124  
QY 59 VRSPNIDQLASHLLFQNAFAQAQVCAVSRVSLTGRPDTRLYDFNSWVRVHAG--NF 116  
Db 125 DEKLLPQLKEAGYTHMVGK-WHLGMYRKECLTRRGFTYCYLLGSDYISHERCTL 183  
QY 117 S-TI-FOYFENGYVTVMSGVKVFHFGISSNHTDDSPYSW-SFPYPHPSSEKYENTKTCRG 173  
Db 184 IDA-LNVTR-CALDFRDGEVATGYKNMYSTNFTKRAIALITN-HPPEKPLFLYLALQS 240  
QY 174 PDGELHANLCPVDVLD--VPEG--TL-PDKOSTEQAIQLEKMKTSASPFFLAVGYHK 227  
Db 241 VHEPLQVPEYLKPY 255  
QY 228 PHIPFRYPKEFKLY 242  
RESULT 15  
ID ARS\_HEMPU STANDARD; PRT; 551 AA.  
AC P14000;  
DT 01-JAN-1990 (Rel. 13, Created)  
DT 01-JAN-1990 (Rel. 13, Last sequence update)  
DT 15-JUL-1998 (Rel. 36, Last annotation update)  
DE ARYL-SULFATASE PRECURSOR (EC 3.1.6.1) (ARYL-SULFATE SULPHOHYDROLASE)  
DE (ARS).  
OS Hemacentrotus pulcherrimus (Sea urchin).  
OC Eukaryota; Metazoa; Echinodermata; Echinozoa; Echinoidea;  
OC Euechinoidea; Echinacea; Echinoida; Strongylocentrotidae;  
OC Hemacentrotus.  
RN [1]  
RP SEQUENCE FROM N.A., AND PARTIAL SEQUENCE.  
RC TISSUE=PLUTEUS;



```

RX MEDLINE; 89030699.
RA SASAKI H., YAMADA K., AKASAKA H., SUZUKI K., SAITO A., SATO M.,
RA SHIMADA H.;
RT *cDNA cloning, nucleotide sequence and expression of the gene for
RT arylsulfatase in the sea urchin (Hemicentrotus pulcherrimus)
RT embryo";
RL Eur. J. Biochem. 177:9-13(1988).
RN [2]
RP SEQUENCE FROM N.A.
RX MEDLINE; 90092130.
RA YAMADA K., AKASAKA K., SHIMADA H.;
RT "Structure of sea-urchin arylsulfatase gene.";
RL Eur. J. Biochem. 186:405-410(1989).
CC -1- FUNCTION: MAY BE A STRUCTURAL COMPONENT OF THE EXTRACELLULAR
CC MATRICES INVOLVED IN CELL MOVEMENT DURING MORPHOGENESIS.
CC -1- CATALYTIC ACTIVITY: A PHENOL SULFATE + H2O = A PHENOL +
CC SULFATE.
CC -1- SUBCELLULAR LOCATION: IN BOTH THE CYTOPLASM AND THE EXTRACELLULAR
CC MATRICES.
CC -1- SIMILARITY: BELONGS TO THE SULFATASE FAMILY.
-----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
-----
DR EMBL; X17015; CAA34881.1; -.
DR PIR; S01793; S01793.
DR PIR; S07089; S07089.
DR HSSP; P15289; LAUK.
DR PROSITE; PS00523; SULFATASE_1; 1.
DR PROSITE; PS00149; SULFATASE_2; 1.
DR PFAM; PF00884; Sulfatase; 1.
DR KW Hydrolase; Signal; Glycoprotein; Extracellular matrix.
FT SIGNAL 1 20
FT CHAIN 21 551 ARYL SULFATASE.
FT MOD_RES 21 21 BLOCKED.
FT MOD_RES 100 100 2-AMINO-3-OXOPROPIONIC ACID (BY
FT SIMILARITY).
FT ACT_SITE 158 158 POTENTIAL.
FT CARBOHYD 164 164 POTENTIAL.
FT CARBOHYD 213 213 POTENTIAL.
FT CARBOHYD 296 296 POTENTIAL.
FT SEQUENCE 551 AA; 60952 MW; 0C7B7505 CRC32;
SQ
Query Match 4.5%; Score 184; DB 1; Length 551;
Best Local Similarity 32.8%; Pred. No. 2.73e-14;
Matches 39; Conservative 28; Mismatches 45; Indels 7; Gaps 6;

Db 53 NVLLVADHMGSGDLTSYGHPTQEQAGFDKMAEGLRFTNGYVGDAVCTPSRSIAIMTGRL 112
|||::: |:: |:: ||:: |:: |:: |:: |:: |:: |:: |:: |:: |:: |:: |::
QY 38 NVLLIIIVDDLRRPS-LGCYGDGKLVSRPNIDQLASHSLLFQNAFAQAQVCAFSRVSELTGR 96

Db 113 PVRIGTGETGVFLPWTKTLGLPKSELITAEAMKEAGYATCMVGK-WHLGINENSSTGGA 170
|::: |::: |::: |::: |::: |::: |::: |::: |::: |::: |::: |::: |::: |:::
QY 97 PDTRTRYL-DFNYSY--W-RVHAGNFS-TIPQYFKENGVTVMGVKVFHPGISNNHTDDSP 150

```

Search completed: Wed Apr 19 21:15:54 2000  
Job time : 395 secs.

\*\*\*\*\*  
Release 3.1A John F. Collins, Biocomputing Research Unit.  
Copyright (c) 1993-1998 University of Edinburgh, U.K.  
Distribution rights by Oxford Molecular Ltd  
\*\*\*\*\*  
MPsrch\_pp protein - protein database search, using Smith-Waterman algorithm  
Run on: Wed Apr 19 20:58:22 2000; MasPar time 35.76 Seconds  
Tabular output not generated.  
Title: >US-09-249-003-2  
Description: (1-550) from US09249003.pep  
Perfect Score: 4069  
Sequence: 1 MPPPRVGRGLLWGLVLSV.....QDHNNYNDQGGDLFQLLMP 550  
Scoring table: PAM 150  
Gap 11  
Searched: 142080 seqs, 47172406 residues  
Post-processing: Minimum Match 0%  
Listing first 45 summaries  
Database: pir62  
1:pir1 2:pir2 3:pir3 4:pir4  
Statistics: Mean 51.177; Variance 98.507; scale 0.520  
Pred. No. is the number of results predicted by chance to have a  
score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.  
SUMMARIES

Result No.	Score	Query Match	Length	ID	Description	Pred. No.
1	4069	100.0	550	1	iduronate-2-sulfatase	0.00e+00
2	3129	76.9	563	2	iduronate-2-sulfatase	0.00e+00
3	1300	31.9	212	2	iduronate-2-sulfatase	1.72e-242
4	282	6.9	497	2	probable sulfatase (E	1.02e-29
5	239	5.9	589	2	arylsulfatase E (EC 3	7.72e-22
6	234	5.8	593	2	arylsulfatase D (EC 3	6.07e-21
7	211	5.2	522	1	N-acetylglucosamine	6.84e-17
8	211	5.2	567	2	arylsulfatase (EC 3.1	6.84e-17
9	213	5.2	583	1	steryl-sulfatase (EC	3.07e-17
10	198	4.9	506	2	cerebroside-sulfatase	1.17e-14
11	195	4.8	507	1	cerebroside-sulfatase	3.80e-14
12	190	4.7	533	1	N-acetylglucosamine	2.65e-13
13	190	4.7	531	2	arylsulfatase (EC 3.1	2.65e-13
14	184	4.5	551	2	arylsulfatase (EC 3.1	2.66e-12
15	182	4.5	970	2	probable sulfatase (E	5.70e-12
16	179	4.4	465	2	probable sulfatase (E	1.78e-11
17	172	4.2	535	1	N-acetylglucosamine	2.47e-10
18	153	3.8	313	1	conserved hypothetical	2.52e-07
19	153	3.8	551	2	arylsulfatase (EC 3.1	2.52e-07
20	136	3.3	787	2	probable sulfatase (E	8.97e-05
21	133	3.3	787	2	probable sulfatase (E	2.44e-04
22	126	3.1	433	1	transcobalamin I prec	2.39e-03
23	126	3.1	571	2	arylsulfatase homolog	2.39e-03

24	126	3.1	919	2	S37786	hypothetical protein	2.39e-03
25	124	3.0	464	2	B35159	arylsulfatase (EC 3.1	4.53e-03
26	120	2.9	190	1	VCBVC	coat protein - cowpea	1.60e-02
27	117	2.9	533	2	S89336	arylsulfatase (EC 3.1	4.04e-02
28	115	2.8	473	2	I54210	arylsulfatase (EC 3.1	7.43e-02
29	114	2.8	540	1	OXECLD	L-aspartate oxidase (	1.00e-01
30	112	2.8	585	1	F64159	hypothetical protein	1.83e-01
31	111	2.7	708	2	B72619	probable NADH dehydro	2.46e-01
32	109	2.7	709	2	T16584	hypothetical protein	4.42e-01
33	105	2.6	44	2	B38075	N-acetylglucosamine	1.40e+00
34	107	2.6	479	2	C75099	hypothetical protein	7.90e-01
35	107	2.6	552	1	KJHUGU	N-acetylglucosamine-6	7.90e-01
36	105	2.6	647	2	C71534	probable transglycola	1.40e+00
37	106	2.6	667	2	C70719	probable carbamoyl-ph	1.05e+00
38	106	2.6	815	2	T15749	hypothetical protein	1.05e+00
39	107	2.6	1034	2	S36758	mgll protein - mouse	7.90e-01
40	107	2.6	4563	1	LPHUB	apolipoprotein B-100	7.90e-01
41	102	2.5	245	2	B33956	tyrosine synthetase	3.24e+00
42	103	2.5	440	2	S73396	hypothetical protein	2.45e+00
43	103	2.5	467	2	E70976	hypothetical protein	2.45e+00
44	102	2.5	518	1	STECCE	glutamate--cysteine	3.24e+00
45	102	2.5	586	1	C64988	hypothetical 67.3 kD	3.24e+00

ALIGNMENTS

RESULT 1  
ENTRY KJHUID #type complete  
TITLE iduronate-2-sulfatase (EC 3.1.1.6.13) precursor - human  
ALTERNATE\_NAMES Chondroitinsulfatase; L-iduronate 2-sulfate 2-sulfohydrolase  
ORGANISM #formal\_name Homo sapiens #common\_name man  
DATE 28-Mar-1991 #sequence\_revision 27-Oct-1995 #text\_change  
24-Sep-1999  
ACCESSIONS A47535; A36483  
REFERENCE A47535  
#authors Wilson, P.J.; Meaney, C.A.; Hopwood, J.J.; Morris, C.P.  
#journal Genomics (1993) 17:773-775  
#title Sequence of the human iduronate 2-sulfatase (IDS) gene.  
#cross-references MUID:94063929  
#accession A47535  
#status translation not shown; translated from GB/EMBL/DBJ  
#molecule\_type DNA  
#residues 1-550 #label W12  
#cross-references GB:LI3329; NID:g405203; PIDN:AAA16877.1; PID:g405205  
REFERENCE A36483  
#authors Wilson, P.J.; Morris, C.P.; Anson, D.S.; Occhiodoro, T.; Bielecki, J.; Clements, P.R.; Hopwood, J.J.  
#journal Proc. Natl. Acad. Sci. U.S.A. (1990) 87:8531-8535  
#title Hunter syndrome: isolation of an iduronate-2-sulfatase cDNA clone and analysis of patient DNA.  
#cross-references MUID:91046030  
#accession A36483  
#molecule\_type mRNA  
#residues 1-550 #label W1L  
#cross-references GB:M58342; GB:M38371; NID:g184561; PIDN:AAA63197.1; PID:g184562  
#note parts of this sequence, including the amino end of the mature protein, were determined by protein sequencing  
#note a form is described with a proteolytic cleavage releasing residue 456 as the amino terminal of the cleaved chain  
GENETICS  
#gene GDB:IDS; SIDS  
#map\_position Xq28-qx28  
#inrons 35/1; 80/3; 140/1; 169/3; 236/3; 293/3; 336/1; 394/1  
#note defects in this gene can cause mucopolysaccharidosis type II, Hunter disease  
FUNCTION  
#description hydrolyzes iduronate-2-sulfate units in dermatan sulfate and heparan sulfate to release sulfate  
CLASSIFICATION #superfamily animal sulfatase  
KEYWORDS glycoprotein; Hunter disease; lysosomal storage disease;

```
lysosome; sulfuric ester hydrolase

FEATURE
1-25      #domain signal sequence #status predicted #label SIG\
26-33     #domain propeptide #status predicted #label PRO\
34-550    #product iduronate-2-sulfatase #status predicted #label
          MAT\
31,115,144,246,280,
325,513,537  #binding_site carbohydrate (Asn) (covalent) #status
              predicted\
84          #modified_site 3-oxoalanine (Cys) #status predicted
SUMMARY     #length 550 #molecular-weight 61872 #checksum 5247
            100.0%; Score 4069; DB 1; Length 550;
Best Local Similarity 100.0%; Pred. No. 0.00e+00;
Matches 550; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Db 1 MPPRTGRLGLWLGLVLSVCVALGSETQANSTTDALNVLIIIVDDLRLPSLCYGDKLYR 60
QY 1 MPPRTGRLGLWLGLVLSVCVALGSETQANSTTDALNVLIIIVDDLRLPSLCYGDKLYR 60
Db 61 SPNIDQLASHLLFQNAFAQAQAVCAPSRVSFLTGRRPDTTRLDFNSYWRVHAGNFSTIP 120
QY 61 SPNIDQLASHLLFQNAFAQAQAVCAPSRVSFLTGRRPDTTRLDFNSYWRVHAGNFSTIP 120
Db 121 QYFKENGVTMSVGKVFHPGIGSSNHTDPSYSWSFPPYHPSSEKYENTKTCRPGDELHA 180
QY 121 QYFKENGVTMSVGKVFHPGIGSSNHTDPSYSWSFPPYHPSSEKYENTKTCRPGDELHA 180
Db 181 NLLCPVDVLDPEGLPKQSTEQAIOLLEKMKTSASPFELAVGYHKPHIPRYPKFQK 240
QY 181 NLLCPVDVLDPEGLPKQSTEQAIOLLEKMKTSASPFELAVGYHKPHIPRYPKFQK 240
Db 241 LYPLENITLAPDPEVPGDGLPPVAYNPWMDIROREDVQALNISVPYGPPIPVDFQKIRQSY 300
QY 241 LYPLENITLAPDPEVPGDGLPPVAYNPWMDIROREDVQALNISVPYGPPIPVDFQKIRQSY 300
Db 301 FASVSYLDTQVGRLLSALDDQLANSTIIAFTSDHGWLGEHGEWAKYSNFDVATHVPLI 360
QY 301 FASVSYLDTQVGRLLSALDDQLANSTIIAFTSDHGWLGEHGEWAKYSNFDVATHVPLI 360
Db 361 FYVPGRTASLPKAGKLPYLDPPDSASQMEPGQSDMLVELVSLFPTLAGLAGLOVPP 420
QY 361 FYVPGRTASLPKAGKLPYLDPPDSASQMEPGQSDMLVELVSLFPTLAGLAGLOVPP 420
Db 421 RCPVPFHVLCREGKLLKHFRLDEEDPYLPCNPRELIAISQYPRPSDIPQWNSDKP 480
QY 421 RCPVPFHVLCREGKLLKHFRLDEEDPYLPCNPRELIAISQYPRPSDIPQWNSDKP 480
Db 481 SLKDIKIMGYSIRTIDYRYTWVVGPNPDEFLANFSDIHAGELYFYVDSPLQDHNMYNDSQ 540
QY 481 SLKDIKIMGYSIRTIDYRYTWVVGPNPDEFLANFSDIHAGELYFYVDSPLQDHNMYNDSQ 540
Db 541 GGDLEFQLLMP 550
QY 541 GGDLEFQLLMP 550

RESULT 2
ENTRY A47153 #type complete
TITLE iduronate-2-sulfatase (EC 3.1.6.13) - mouse
ORGANISM #formal_name Mus musculus #common_name house mouse
DATE 03-May-1994 #sequence_revision 03-May-1994 #text_change
22-Jun-1999
ACCESSIONS A47153
REFERENCE
#authors Daniele, A.; Faust, C.J.; Herman, G.E.; Di Natale, P.;
Ballabio, A.
#journal Genomics (1993) 16:755-757
#title Cloning and characterization of the cDNA for the murine
iduronate sulfatase gene.
#cross-references MUID:93315172
#accession A47153
#status preliminary
```

```
##molecule_type mRNA
##residues 1-563 ##label DAN
##cross-references GB:U07921; NID:g349713; PIDN:AAA37880.1; PID:g349714
CLASSIFICATION #superfamily animal sulfatase
KEYWORDS sulfuric ester hydrolase
FEATURE
99 #modified_site 3-oxoalanine (Cys) #status predicted
SUMMARY #length 563 #molecular-weight 63437 #checksum 8505

Query Match 76.9%; Score 3129; DB 2; Length 563;
Best Local Similarity 79.0%; Pred. No. 0.00e+00;
Matches 433; Conservative 58; Mismatches 53; Indels 4; Gaps 4;

Db 19 ARAIWRQLSFLLSGFCIALESAAQNSATDALNILLIIVDDLRLPSLCYGDKLVSPN 78
QY 4 PRTRGRLGLWLGLVLSVCVALGSETQANSTTDALNVLIIIVDDLRLPSLCYGDKLVSPN 63
Db 79 IDOLASHSVLFQNAFAQAQAVCAPSRVSFLTGRRPDTTRLDFNSYWRVHAGNFSTIPQYF 138
QY 64 IDOLASHSVLFQNAFAQAQAVCAPSRVSFLTGRRPDTTRLDFNSYWRVHAGNFSTIPQYF 123
Db 139 KENGYVTMSVGKVFHPGIGSSNHTDPSYSWSFPPYHPSSEKYENTKTCRQDGLHANLL 198
QY 124 KENGYVTMSVGKVFHPGIGSSNHTDPSYSWSFPPYHPSSEKYENTKTCRQDGLHANLL 183
Db 199 CPVDVADVPETLPDRKQSTEQAIOLLEKMKTSASPFELAVGYHKPHIPRYPKFQKLYP 258
QY 184 CPVDVADVPETLPDRKQSTEQAIOLLEKMKTSASPFELAVGYHKPHIPRYPKFQKLYP 243
Db 259 LENITLAPDPEVPGDGLPPVAYNPWMDIROREDVQALNISVPYGPPIPVDFQKIRQSYFAS 318
QY 244 LENITLAPDPEVPGDGLPPVAYNPWMDIROREDVQALNISVPYGPPIPVDFQKIRQSYFAS 303
Db 319 VSYLDTQVGRLLSALDDQLANSTIIAFTSDHGWLGEHGEWAKYSNFDVATHVPLIYV 378
QY 304 VSYLDTQVGRLLSALDDQLANSTIIAFTSDHGWLGEHGEWAKYSNFDVATHVPLIYV 363
Db 379 PGRTAPLPAAGKLPYLDPPDSASQMEPGQSDMLVELVSLFPTLAGLAGLOVPP 438
QY 364 PGRTAPLPAAGKLPYLDPPDSASQMEPGQSDMLVELVSLFPTLAGLAGLOVPP 423
Db 439 SLLFMLSFAEKAR-IFRSIC-SSMTWKRQTCCLVPRELIAISQYPRPADPQWNSDKP 496
QY 424 VPSFHVLCREGKLLKHFRLDE-EDPYLPCNPRELIAISQYPRPSDIPQWNSDKP 482
Db 497 NDIRSHWDIYTH-VDYRYTWVVGDPSEFLANFSDIHAGELYFYVDSPLQDHNMYNDSQ 555
QY 483 KDIKIMGYSIRTIDYRYTWVVGPNPDEFLANFSDIHAGELYFYVDSPLQDHNMYNDSQ 542
Db 556 GLLHSLRP 563
QY 543 DLFQLLMP 550

RESULT 3
ENTRY PNO565 #type fragment
TITLE iduronate-2-sulfatase (EC 3.1.6.13) (clone MTA13) - mouse
(fragment)
ALTERNATE_NAMES iduronate sulfatase
ORGANISM #formal_name Mus musculus #common_name house mouse
DATE 05-Aug-1994 #sequence_revision 05-Aug-1994 #text_change
17-Mar-1999
ACCESSIONS PNO565
REFERENCE
#authors Daniele, A.; Russo, T.; Ballabio, A.; Di Natale, P.;
#journal Biochem. Biophys. Res. Commun. (1993) 194:1030-1037
#title The mouse iduronate sulfatase gene: Identification of a novel
transcript.
#cross-references MUID:93356764
#accession PNO565
##molecule_type mRNA
##residues 1-212 ##label DAN
##note the authors translated the codon CAT for residue 23 as
```



4

D	b	113	SVCPTSPSAIVTGRQPIRTGVGEERIFLPWTWTTGLPLVEYVETIAEAMKGGYTTCGMVGK- 171 :  :    :  :    :  :    :  :    :  :    :  :    :  :    :  :    :  :
Q	y	82	AVCAPSRVSFLTGRPPDTRLTY-DENSY--WRVHA-GNFS-TIPQYFKENGYYTWTSVGKV 136
D	b	172	WHGINENSSSDGA 185
Q	y	137	FHPGISSNHDTDSP 150    : : :: :
R	E	S	U
RESULT	9		
ENTRY	KJHUAC	#type complete	
TITLE	steryl-sulfatase (EC 3.1.6.2) precursor - human		
ALTERNATE_NAMES	arylsulfatase C; steroid sulfatase (STS); steryl-sulfate sulfohydrolase; sterylsulfatase		
ORGANISM	#formal name Homo sapiens #common_name man		
DATE	21-May-1990 #sequence_revision 27-Oct-1995 #text_change 20-Aug-1999		
ACCSSIONS	A32641; A57116; A25961; S05415; S05423; I52800; I65619		
REFERENCE	A32641		
#authors	Stein, C.; Hille, A.; Seidel, J.; Rijnhout, S.; Waheed, A.; Schmidt, B.; Geuze, H.; von Figura, K.		
#journal	J. Biol. Chem. (1989) 264:13865-13872		
#title	Cloning and expression of human steroid-sulfatase. Membrane topology, glycosylation, and subcellular distribution in BHK-21 cells.		
#cross-references	MUID:89340479		
#accession	A32641		
#molecule_type	mRNA		
#residues	1-583 ##label STE		
#cross-references	GB:J04964; NID:g338564; PIDN:AAA60597.1; PID:g338565		
#experimental_source	BHK-21 cells		
#note	parts of this sequence were determined by protein sequencing		
REFERENCE	A57116		
#authors	Yen, P.H.		
#citation	unpublished results 1988, cited by GenBank		
#accession	A57116		
#molecule_type	mRNA		
#residues	1-22,'E','24'-583 ##label YEN1		
#cross-references	GB:M16505; NID:g338513; PIDN:AAA60596.1; PID:g338514		
REFERENCE	A25961		
#authors	Yen, P.H.; Allen, E.; Marsh, B.; Mohandas, T.; Wang, N.; Taggart, R.T.; Shapiro, L.J.		
#journal	Cell (1987) 49:443-454		
#title	Cloning and expression of steroid sulfatase cDNA and the frequent occurrence of deletions in STS deficiency: implications for X-Y interchange.		
#cross-references	MUID:87187642		
#accession	A25961		
#molecule_type	mRNA		
#residues	1-22,'E','24'-456, 'LRTHTPSRGPSSPPTSTFWQRALPHTCASVSGVMSPITTHLYSLIPP', 'KIPERE' ##label YEN2		
#cross-references	GB:M16505; NID:g338513		
#note	this sequence revised in A57116		
#note	part of this sequence, including the amino end of the mature protein, was determined by protein sequencing		
REFERENCE	S05414		
#authors	Kawano, J.I.; Kotani, T.; Ohtaki, S.; Minamino, N.; Matsuo, H.; Oinuma, T.; Aikawa, E.		
#journal	Biochim. Biophys. Acta (1989) 997:199-205		
#title	Characterization of rat and human steroid sulfatases.		
#cross-references	MUID:89352671		
#accession	S05415		
#molecule_type	protein		
#residues	22-43,'X','45'-46 ##label KAW		
REFERENCE	S05423		
#authors	Dibbelt, L.; Otto, J.; Kuss, F.		
#journal	Biol. Chem. Hoppe-Seyler (1989) 370:847-848		
#title	The N-terminal amino-acid sequence of human placental sterylsulfatase.		
#cross-references	MUID:90074181		

```

#accession      S05423
#molecule_type protein
#residues      22-35 ##label DIB
REFERENCE
I52800
#authors        Yen, P.H.; Marsh, B.; Allen, E.; Tsai, S.P.; Ellison, J.;
                  Connolly, L.; Neiswanger, K.; Shapiro, L.J.
#journal        Cell (1988) 55:1123-1135
#title          The human X-linked testis sulfatase gene and a Y-encoded
                  pseudogene: evidence for an inversion of the Y chromosome
                  during primate evolution.
#cross-references MIMD:89077541
#accession      I52800
#status         translated from GB/EMBL/DBJ
#molecule_type DNA
#residues      134-274 ##label YEN3
#cross-references GB:IM23945; NID:g338604; PIDN:AAA60598.1; PID:g338607
#accession      I65619
#status         translated from GB/EMBL/DBJ
#molecule_type DNA
#residues      461-583 ##label YEN4
#cross-references GB:IM23556; NID:g338605; PIDN:AAA60599.1; PID:g338608
GENETICS
#gene           GDR:STS

```

```

GENETICS
#gcross_references GB:W2353J6; MID:G9386005; FIDN:AA800595.1; FID:G9386005
#gene GDB:STS
##cross-references GDB:120393; OMIM:308100
#map_position Xp22.32-Xp22.32
#note defects in this gene can cause X-linked ichthyosis
FUNCTION
#description hydrolyzes 3beta-hydroxysteroid sulfates to release sulfate
CLASSIFICATION #superfamily animal sulfatase
KEYWORDS endoplasmic reticulum; glycoprotein; lysosome; microsome;
sulfuric ester hydrolase; transmembrane protein
FEATURE
1-21 #domain signal sequence #status predicted #label SIG\
22-503 #product steryl-sulfatase #status predicted #label MAT\
185-211 #domain transmembrane #status predicted #label TM1\
213-237 #domain transmembrane #status predicted #label TM2\
47,259 #binding_site carbohydrate (Asn) (covalent) #status
experimental\
75 #modified_site 3-oxoalanine (Cys) #status predicted\
333,459 #binding_site carbohydrate (Asn) (covalent) #status
absent
SUMMARY
#length 583 #molecular-weight 65492 #checksum 7062

Query Match 5.2%; Score 213; DB 1; Length 583;
Best Local Similarity 27.6%;
Matches 92; Conservative 85; Mismatches 127; Indels 29; Gaps 24;

```

[illegible]

10  
RESULT

```
ENTRY
TITLE      A54190      #type complete
ALTERNATE_NAMES
ORGANISM   cerebroside-sulfatase (EC 3.1.6.8) precursor - mouse
DATE       18-Aug-1995 #sequence_revision 18-Aug-1995 #text_change
29-Sep-1999

ACCESSIONS
REFERENCE  A54190; S34293; A38075; S34294
#authors   Kreyling, J.; Polten, A.; Hess, B.; von Figura, K.; Menz, K.;
            Steiner, F.; Gieselmann, V.
#journal   Genomics (1994) 19:249-256
#title     Structure of the mouse arylsulfatase A gene and cDNA.
#cross-references MUID:94245194
#accession A54190
#status    preliminary
#molecule_type mRNA
#residues  1-506 #label KRE
#cross-references GB:X73230; NID:g312271; PIDN:CAA51702.1; PID:g312272
S34293
#authors   Kreyling, J.; Polten, A.; Hess, B.; Menz, K.; Steiner, F.;
            Gieselmann, V.
#submission submitted to the EMBL Data Library, June 1993
#accession S34293
#status    preliminary
#molecule_type DNA
#residues  1-506 #label KR2
#cross-references EMBL:X73231; NID:g312268; PIDN:CAA51703.1;
            PID:g312269

REFERENCE  A38075
#authors   Grompe, M.; Pieretti, M.; Caskey, C.T.; Ballabio, A.
#journal   Genomics (1992) 12:755-760
#title     The sulfatase gene family: cross-species PCR cloning using
            the MOPAC technique.
#cross-references MUID:92241876
#accession A38075
#molecule_type mRNA
#residues  27-72 #label GRO
#cross-references GB:M82876; NID:gl92027; PIDN:AAA37260.1; PID:gl92028
#note      sequence extracted from NCBI backbone (NCBIN:98744,
            NCBI:P:98749)
#note      the protein sequence from Fig. 3 includes conserved
            regions used in the primer
CLASSIFICATION #superfamily animal sulfatase
KEYWORDS       glycoprotein; sulfuric ester hydrolase
FEATURE
1-17           #domain signal sequence #status predicted #label SIG\
18-506         #product cerebroside-sulfatase #status predicted #label
            MAT\
68             #modified_site 3-oxoalanine (Cys) #status predicted\
157,183,349    #binding_site carbohydrate (Asn) (covalent) #status
            predicted
SUMMARY        #length 506 #molecular-weight 53776 #checksum 7525

Query Match      4.9%; Score 198; DB 2; Length 506;
Best Local Similarity 42.5%; Pred. No. 1.17e-14;
Matches 31; Conservative 16; Mismatches 25; Indels 1; Gaps 1;

Db 15 STASPPNLLIFADDLGVGDLSGTHSPSTPNLDQLEGGRLFTDFVVPVSLCTPSRAA 74
   ||: : l: |||: ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Qy 32 STTDALNVLIIIVDLR-PSLIGCYGDKLVRSNPIDQLASHSLLFQNAFAQQAVCAPSRVS 90
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 75 LLTGRLPVRSAMY 87
Qy 91 FLTGRRPDPTRLY 103

RESULT 11
ENTRY KJHUA
TITLE cerebroside-sulfatase (EC 3.1.6.8) precursor - human
ALTERNATE_NAMES arylsulfatase A (ASA); arylsulfatase (EC 3.1.6.1)
                (misidentification)
ORGANISM #formal_name Homo sapiens #common_name man
DATE 21-Nov-1993 #sequence_revision 27-Oct-1995 #text_change

11-Jun-1999
ACCESSIONS S11031; G02857; A32207; S23932
REFERENCE S11031
#authors   Kreyling, J.; von Figura, K.; Gieselmann, V.
#journal   Eur. J. Biochem. (1990) 191:627-631
#title     Structure of the arylsulfatase A gene.
#cross-references MUID:90361046
#accession S11031
#molecule_type DNA
#residues  1-507 #label KRE
#cross-references EMBL:X52150; NID:g28859; PIDN:CAA36398.1; PID:g28860
H01749
#authors   Adams, M.D.; Kerlavage, A.R.; Fuldner, R.A.; Phillips, C.A.;
            Venter, J.C.
#submission submitted to the EMBL Data Library, June 1996
#accession G02857
#status    translated from GB/EMBL/DBDJ
#molecule_type DNA
#residues  'MS', 1-507 #label ADA
#cross-references EMBL:U62317; NID:gl39959; PID:gl3
#note      an incorrect initiation codon was used
REFERENCE  A32207
#authors   Steinh, C.; Gieselmann, V.; Kreyling, J.; Schmidt, B.;
            Pohlmann, K.; Waheed, A.; Meyer, H.E.; O'Brien, J.S.; von
            Figura, K.
#journal   J. Biol. Chem. (1989) 264:1252-1259
#title     Cloning and expression of human arylsulfatase A.
#cross-references MUID:89093115
#accession A32207
#molecule_type mRNA
#residues  1-358, 'RPPAAGHRQPSAVSLLLPVLPKRGPGWGCADKWQGSLLHP',
            402-507 #label STE
#cross-references GB:X52151; GB:J04442; GB:J04593
#note      parts of this sequence, including the amino end of the
            mature protein, were determined by protein sequencing
REFERENCE  S23932
#authors   Fujii, T.; Kobayashi, T.; Honke, K.; Gasa, S.; Ishikawa, M.;
            Shimizu, T.; Makita, A.
#journal   Blochim. Biophys. Acta (1992) 1122:93-98
#title     Protolytic processing of human lysosomal arylsulfatase A.
#cross-references MUID:92338230
#accession S23932
#molecule_type protein
#residues  20-29, 31-33; 434-479 #label FUJ
#experimental_source placenta
REFERENCE  A57113
#authors   Schmidt, B.; Selmer, T.; Ingendoh, A.; von Figura, K.
#journal   Cell (1995) 82:271-278
#title     A novel amino acid modification in sulfatases that is
            defective in multiple sulfatase deficiency.
#cross-references MUID:95354208
#contents   annotation; identification of 3-oxoalanine,
            2-amino-3-oxopropanoic acid

GENETICS
#gene      GDB:ARSA
#cross-references GDB:119007; OMIM:250100
#map_position 22q13.31-22qter
#introns    73/2; 153/3; 226/3; 283/2; 325/1; 367/3; 402/1
#note       defects in this gene can cause metachromatic leukodystrophy
FUNCTION    hydrolyzes cerebroside 3-sulfate to release sulfate; can also
            hydrolyze galactose-3-sulfate, ascorbate 2-sulfate, and
            many phenol sulfates
            #superfamily animal sulfatase
            glycoprotein; lysosomal storage disease; lysosome; sulfuric
            ester hydrolase
CLASSIFICATION
KEYWORDS
FEATURE
1-19         #domain signal sequence #status predicted #label SIG\
20-507       #product cerebroside-sulfatase #status predicted #label
            MAT\
20-444,445-507 #product cerebroside-sulfatase component b #status
            predicted #label MCB\
20-444,448-507 #product cerebroside-sulfatase component c #status
```



20-444,449-507	predicted #label MCC #product_cerebroside-sulfatase minor component c #status
69	predicted #label MCD #modified_site 3-oxoalanine (Cys) #status experimental\
158	#binding_site carbohydrate (Asn) (covalent) #status experimental\
184,350	#binding_site carbohydrate (Asn) (covalent) #status predicted
SUMMARY	#length 507 #molecular-weight 53588 #checksum 6235 Query Match 4.8%; Score 195; DB 1; Length 507; Best_Local Similarity 43.3%; Pred. No. 3.80e-14; Matches 29; Conservative 15; Mismatches 22; Indels 1; Gaps 1;
Db	22 NVILFADLDGDLGCGHPSSPTPNLDQLAAGRGFTDFYVPVSLCTPSRAALLTGR 81 I::I:: I::I:: I::I::I::I:: I::I::I::I:: I::I::I::I::
Qy	38 NVLLIIVDDL-PSLGGYGDKLVRSPNIDQLASHLLFNQAEQAQVCAPSRVSLTGR 96 I::I:: I::I:: I::I::I::I:: I::I::I::I:: I::I::I::I::
Db	82 PVRMGMY 88 I::I:: I::I:: I::I::I::I:: I::I::I::I:: I::I::I::I::
Qy	97 PDTRRLY 103 I::I:: I::I:: I::I::I::I:: I::I::I::I:: I::I::I::I::
RESULT	12
ENTRY	KJHUAB #type complete
TITLE	N-acetylgalactosamine-4-sulfatase (EC 3.1.6.12) precursor - human
ALTERNATE_NAMES	arylsulfatase B (ASB); chondroitinase; chondroitinsulfatase; GAS; N-acetyl-D-galactosamine-4-sulfate 4-sulfohydrolase
ORGANISM	#formal_name Homo sapiens #common_name man
DATE	31-Dec-1993 #sequence_revision 27-Oct-1995 #text_change 11-Jun-1999
ACCESSIONS	S35990; S33307; A35078; A45659; A24449; B42449; C42449; IS4217; A56865
REFERENCE	S35990 Peters, C.W.B. #authors #submission submitted to the EMBL Data Library, March 1993 #accession S35990
	#molecule_type DNA #residues 1-533 #label PET #cross-references EMBL:X72735; NID:g289009; PIDN:CAA51272.1; PID:g825628
REFERENCE	S33307 #authors Modaresi, S.; Rupp, K.; von Figura, K.; Peters, C. #journal Biol. Chem. Hoppe-Seyler (1993) 374:327-335 #title Structure of the human arylsulfatase B gene. #cross-references MUID:93332648 #accession S33307
	#molecule_type DNA #residues 1-104 #label MOD #cross-references EMBL:X72735; EMBL:X72736; EMBL:X72737; EMBL:X72738; EMBL:X72739; EMBL:X72740; EMBL:X72741; EMBL:X72742 the enzyme is referred to as EC 3.1.6.9
REFERENCE	A35078 #authors Peters, C.; Schmidt, B.; Rommerskirch, W.; Rupp, K.; Zuehlisdorff, M.; Vingron, M.; Meyer, H.E.; Pohlmann, R.; von Figura, K. #journal J. Biol. Chem. (1990) 265:3374-3381 #title Phylogenetic conservation of arylsulfatases. cDNA cloning and expression of human arylsulfatase B. #cross-references MUID:90153994 #accession A35078
	#molecule_type mRNA #residues 1-357, 'V', 359-533 #label PE2 #cross-references GB:J05225; NID:g179076; PIDN:AAA51784.1; PID:g179077 #note parts of this sequence were determined by protein sequencing the enzyme is referred to as EC 3.1.6.1
REFERENCE	A45659 #authors Litjens, T.; Morris, C.P.; Gibson, G.J.; Beckmann, K.R.; Hopwood, J.J. #journal Biochem. Int. (1991) 24:209-215 #title Human N-acetylgalactosamine-4-sulphatase: protein maturation



DATE 17-Jul-1998 #sequence\_revision 17-Jul-1998 #text\_change 24-Sep-1999  
ACCESSIONS E70533  
REFERENCE A70500  
#authors Cole, S.T.; Brosch, R.; Parkhill, J.; Garnier, T.; Churcher, C.; Harris, D.; Gordon, S.V.; Eiglmeier, K.; Gas, S.; Barry III, C.E.; Tekala, F.; Badcock, K.; Basham, D.; Brown, D.; Chillingworth, T.; Connor, R.; Davies, R.; Devlin, K.; Feltwell, T.; Gentles, S.; Hamlin, N.; Holroyd, S.; Hornsby, T.; Jagels, K.; Krogh, A.; McLean, J.; Moule, S.; Murphy, L.; Oliver, S.; Osborne, J.; Quail, M.A.; Rajandream, M.A.; Rogers, J.; Rutter, S.; Seeger, K.; Skelton, S.; Squares, S.; Squires, R.; Sulston, J.E.; Taylor, K.; Whitehead, S.; Barrell, B.G.  
#journal Nature (1998) 393:537-544  
#title Deciphering the biology of Mycobacterium tuberculosis from the complete genome sequence.  
#cross-references MUID:98295987  
#accession E70533  
#status preliminary; nucleic acid sequence not shown; translation not shown  
#molecule\_type DNA  
#residues 1-970 #label COL  
#cross-references GB:Z96070; GB:AL123456; NID:g3261791; PID:e320915; PID:g2181956  
#experimental\_source strain H37Rv  
GENETICS  
#gene atsB  
KEYWORDS sulfuric ester hydrolase  
FEATURE  
258  
SUMMARY  
#length 970 #molecular-weight 105679 #status predicted  
#modified\_site 3-oxoalanine (Cys)  
Query Match 4.5%; Score 182; DB 2; Length 970;  
Best Local Similarity 30.0%; Pred. No. 5.70e-12;  
Matches 39; Conservative 39; Mismatches 40; Indels 12; Gaps 9;  
Db 210 DAPNVLIVLIDDAGFGPDTFGGA-IRPTLSRLAQNGLIY-NRFHVTAVCSPTRAALLT 267  
Qy 35 DALNVLLIIIVDDLRL-PSLGCYGDKLVRSPNIDQLASHSLLFQNAFAQAVCAPSRVSLT 93  
Db 268 GRNHRVGFSGVCEPFGPYGYSAVRPSRCAALPRILRDNGYVTGAFGR-WHLTPDNVQG 326  
Qy 94 GR--RPDTRLYDF-NSY--WR-VHAGNFSTIPQTKENGIVTMSVGKVEHPGISSNHT 146  
Db 327 AAGPFDNWPL 336  
Qy 147 DDSPY-SWSF 155

Search completed: Wed Apr 19 21:09:02 2000  
Job time : 640 secs.

\*\*\*\*\*

(TM)

\*\*\*\*\*

Release 3.1A John F. Collins, Biocomputing Research Unit.  
 Copyright (c) 1993-1998 University of Edinburgh, U.K.  
 Distribution rights by Oxford Molecular Ltd

MPsrch\_pp protein - protein database search, using Smith-Waterman algorithm

Run on: Wed Apr 19 20:52:10 2000; MasPar time 16.52 Seconds  
 Tabular output not generated. 788.664 Million cell updates/sec

Title: >US-09-249-003-2

Description: (1-550) from US09249003.pep

Perfect Score: 4069  
 Sequence: 1 MPPTPTGRGLLWGLVLSSV.....QDHNMYNDSOGDLFQLLMP 550

Scoring table: PAM 150  
 Gap 11

Searched: 188963 seqs, 23686106 residues

Post-processing: Minimum Match 0%  
 Listing first 45 summaries

Database: a-geneseq36  
 1:genesep

Statistics: Mean 37.057; Variance 165.341; scale 0.224

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description	Pred. No.
1	4069	100.0	550	1 W69786	Human iduronate-2-sulf	0.00e+00
2	4069	100.0	550	1 W47298	Human IDS.	0.00e+00
3	173	4.3	406	1 Y02687	Human secreted protein	1.44e-04
4	176	4.3	502	1 W90233	Human sulphamidase pro	8.28e-05
5	172	4.2	556	1 R51355	Sequence of murine bon	1.73e-04
6	164	4.0	86	1 Y12260	Human 5' ESR secreted	7.40e-04
7	136	3.3	25	1 W47300	Human IDS peptide frag	1.05e-01
8	136	3.3	25	1 W69785	Human iduronate-2-sulf	1.05e-01
9	109	2.7	226	1 W41995	Flea cysteine protease	9.52e+00
10	109	2.7	327	1 W41994	Flea cysteine protease	9.52e+00
11	109	2.7	327	1 W41993	Flea cysteine protease	9.52e+00
12	110	2.7	1671	1 R06341	Signal peptide and fir	8.10e+00
13	110	2.7	2721	1 P70647	Sequence of N-terminal	8.10e+00
14	107	2.6	351	1 R44211	Caffeine demethylase.	1.31e+01
15	107	2.6	4536	1 W96826	Amino acid sequence of	1.31e+01
16	107	2.6	4536	1 W41262	Apolipoprotein B-100.	1.31e+01
17	102	2.5	197	1 W81110	Canine Fc-epsilon-rece	2.89e+01
18	102	2.5	197	1 W81115	Canine Fc-epsilon-rece	2.89e+01
19	102	2.5	197	1 W81109	Canine Fc-epsilon-R (1	2.89e+01
20	102	2.5	253	1 W81113	Canine Fc-epsilon-rece	2.89e+01
21	101	2.5	353	1 W88346	Salmonella enterica O	3.37e+01
22	100	2.5	710	1 R75714	Eph-related PTK Cek7.	3.94e+01
23	100	2.5	722	1 R75705	Eph-related PTK Cek7.	3.94e+01

Eph-related PTK Cek7+. 3.94e+01  
 Canine Fc-epsilon-rece 4.60e+01  
 Marek's disease virus 7.30e+01  
 Acid alpha-amylase enz 6.26e+01  
 Pro-C5 polypeptide. 7.30e+01  
 Human iduronate-2-sulf 8.50e+01  
 Human IDS peptide frag 8.50e+01  
 Human derived light ch 1.15e+02  
 Human-derived RN3 phag 1.15e+02  
 C6 human sfv antibody 1.34e+02  
 C6.5/218 single chain 1.34e+02  
 Specific co-activator 9.89e+01  
 DPM2 mannosyl transfer 1.34e+02  
 Pacific yew taxadiene 1.15e+02  
 HSV-2 strain SB5 Conti 1.80e+02  
 HSV-2 ISP35 protein. 1.80e+02  
 Epic protein. 1.80e+02  
 Paired basic amino aci 1.80e+02  
 HSV-2 protease, ICP35. 1.80e+02  
 HSV-2 protease. 1.80e+02  
 HSV-2 strain SB5 Conti 1.80e+02  
 N-tera 2D1 autotaxin p 1.55e+02

## ALIGNMENTS

## RESULT 1

ID W69786 standard; Protein; 550 AA.  
 AC W69786;  
 DT 03-NOV-1998 (first entry)  
 DE Human iduronate-2-sulfatase.  
 KW Human; iduronate-2-sulfatase; IDS; liver; glycosylated; heparin sulphate;  
 KW dermatan sulphate; lysosome; Hunter syndrome; severe mental retardation;  
 KW skeletal deformity; stiff joint.  
 OS Homo sapiens.  
 PN US5798239-A.  
 PD 25-AUG-1998.  
 PF 07-JUN-1995; 484494.  
 PR 17-DEC-1992; US-991973.  
 PR 12-NOV-1991; US-790362.  
 PR 28-NOV-1994; US-345212.  
 PR 07-JUN-1995; US-484494.  
 PA (WOMEN-) WOMEN'S & CHILDREN'S HOSPITAL.  
 PI Anson DS, Bielicki J, Clements PR, Hopwood JJ, Morris CP,  
 PI Occhiodoro T, Wilson PJ;  
 DR WPI; 98-480382/41.  
 DR N-PSDB: V52836.  
 PT Production of glycosylated iduronate-2-sulphatase enzyme (IDS) -  
 PT comprises culturing a host cell containing a nucleic acid encoding  
 PT the enzymatically active iduronate-2-sulphatase polypeptide  
 PT Example 1; Fig 1; 53pp; English.  
 CC A method has been developed for the production of glycosylated  
 CC iduronate-2-sulphatase enzyme (IDS). The method comprises culturing  
 CC a host cell containing a nucleic acid encoding the enzymatically  
 CC active IDS polypeptide where the host cell glycosylates the polypeptide  
 CC to a greater degree than a native IDS polypeptide expressed by a  
 CC natural human liver cell. The present sequence represents human  
 CC IDS which is used in an example from the present invention. The  
 CC recombinant IDS is used to treat IDS deficiency where heparin sulphate  
 CC and dermatan sulphate accumulates in lysosomes resulting in Hunter  
 CC syndrome which is manifested by e.g. severe mental retardation,  
 CC skeletal deformities and stiff joints. The recombinant IDS possesses  
 CC inter alia an improved half life inter alia and improved uptake  
 CC properties in comparison to the naturally glycosylated molecule.  
 SQ Sequence 550 AA;

Query Match 100.0%; Score 4069; DB 1; Length 550;  
 Best Local Similarity 100.0%; Pred. No. 0.00e+00;  
 Matches 550; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 DB 1 MPPTPTGRGLLWGLVLSSVCVVGSTQANSTTDALNVLIIIVDDLRPSLCGYGDKLVR 60  
 QY 1 MPPTPTGRGLLWGLVLSSVCVVGSTQANSTTDALNVLIIIVDDLRPSLCGYGDKLVR 60

Db 61 SPNIDQASHLLFQNAFAQAQAVCAPSRVSFLTGRPPDTTRLYDFNSYWRVHAGNFSTIP 120  
 QY 61 SPNIDQASHLLFQNAFAQAQAVCAPSRVSFLTGRPPDTTRLYDFNSYWRVHAGNFSTIP 120  
 Db 121 QYFKENGVTMSVGVKVFHPGIGSSNHTDDSPYSWSFPPYPHPSSSEKYENTKTCRGPDELHA 180  
 QY 121 QYFKENGVTMSVGVKVFHPGIGSSNHTDDSPYSWSFPPYPHPSSSEKYENTKTCRGPDELHA 180  
 Db 181 NLLCPVDVLDVPEGTLPDKQSTEQAIQLLEKMKTSASPFLLAVGYHKPHIPRYPKFQK 240  
 QY 181 NLLCPVDVLDVPEGTLPDKQSTEQAIQLLEKMKTSASPFLLAVGYHKPHIPRYPKFQK 240  
 Db 241 LYPLENITLAPDPEVPGDGLPPVAYNPWMDIRQEDVQALNLSVPGPLPVDFQKRIQSY 300  
 QY 241 LYPLENITLAPDPEVPGDGLPPVAYNPWMDIRQEDVQALNLSVPGPLPVDFQKRIQSY 300  
 Db 301 FASVSYLDTQVGRLLSALDDLOLANSTIIAFTSDHGALGHEGWAKYSNFDVATHVPLI 360  
 QY 301 FASVSYLDTQVGRLLSALDDLOLANSTIIAFTSDHGALGHEGWAKYSNFDVATHVPLI 360  
 Db 361 FYVPGRTASLPEAGEKLPFYLDPPDSASQLMEPGRQSDMLVELVSLPPTLAGLAGLOVPP 420  
 QY 361 FYVPGRTASLPEAGEKLPFYLDPPDSASQLMEPGRQSDMLVELVSLPPTLAGLAGLOVPP 420  
 Db 421 RCPVPSFHVLCRGGKLLKHFRFRDLEEDPYLPGNPRELIAYSQYPRPSDIPQWNSDKP 480  
 QY 421 RCPVPSFHVLCRGGKLLKHFRFRDLEEDPYLPGNPRELIAYSQYPRPSDIPQWNSDKP 480  
 Db 481 SLKDIKIMGYSIRTIIDYRYTVWVGFNDFEFLANFSDIHAGELYFVDSPLQDHNMYNDSQ 540  
 QY 481 SLKDIKIMGYSIRTIIDYRYTVWVGFNDFEFLANFSDIHAGELYFVDSPLQDHNMYNDSQ 540  
 Db 541 GGDLFQLLMP 550  
 QY 541 GGDLFQLLMP 550

## RESULT 2

ID W47298 standard; Protein; 550 AA.  
 AC W47298;  
 DT 01-JUN-1998 (first entry)  
 DE Human IDS.  
 KW Human; iduronate 2-sulphatase; IDS; treatment;  
 KW Hunter syndrome.  
 OS Homo sapiens.  
 PN US5728381-A.  
 PD 17-MAR-1998.  
 PF 07-JUN-1995; 484493.  
 PR 17-DEC-1992; US-991973.  
 PR 12-NOV-1991; US-790362.  
 PR 28-NOV-1994; US-345212.  
 PR 07-JUN-1995; US-484493.  
 PA (ANSO/) ANSON D S.  
 PA (BIEL/) BIELICKI J.  
 PA (CLEM/) CLEMENTS P R.  
 PA (HOPW/) HOPWOOD J J.  
 PA (MORR/) MORRIS C P.  
 PA (OCCH/) OCCHIODORO T.  
 PA (WILS/) WILSON P J.  
 PI Anson DS, Bielicki J, Clements PR, Hopwood JJ, Morris CP,  
 PI Occhiodoro T, Wilson PJ;  
 DR WPI: 98-206530/18.  
 DR N-PSDB; V15698.  
 PT Treatment of iduronate 2-sulphatase deficiency - comprises  
 PT administering recombinant iduronate 2-sulphatase  
 PS Claim 6; Columns 23-28; 53pp; English.  
 CC The present sequence is human iduronate 2-sulphatase (IDS).  
 CC IDS deficiency can be treated by administering a recombinant human  
 CC IDS that is more highly glycosylated than the naturally occurring  
 CC enzyme, useful in the treatment of Hunter syndrome. The recombinant  
 CC IDS may be administered in 0.5 microg/kg to 20 mg/kg doses. The  
 CC administration route is oral, intravenous, intraperitoneal,

CC intramuscular, subcutaneous or intranasal. The recombinant IDS has  
 CC better uptake properties and/or a longer half-life in vivo, and is  
 CC thus more efficient than naturally glycosylated IDS.  
 SQ Sequence 550 AA;  
 Query Match 100.0%; Score 4069; DB 1; Length 550;  
 Best Local Similarity 100.0%; Pred. No. 0.00e+00;  
 Matches 550; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 Db 1 MPPPTRGLLMWGLVLSVVCVALGSETQANSTTDALNVLIIIVDDRSLGCGYDKLVR 60  
 QY 1 MPPPTRGLLMWGLVLSVVCVALGSETQANSTTDALNVLIIIVDDRSLGCGYDKLVR 60  
 Db 61 SPNIDQASHLLFQNAFAQAQAVCAPSRVSFLTGRPPDTTRLYDFNSYWRVHAGNFSTIP 120  
 QY 61 SPNIDQASHLLFQNAFAQAQAVCAPSRVSFLTGRPPDTTRLYDFNSYWRVHAGNFSTIP 120  
 Db 121 QYFKENGVTMSVGVKVFHPGIGSSNHTDDSPYSWSFPPYPHPSSSEKYENTKTCRGPDELHA 180  
 QY 121 QYFKENGVTMSVGVKVFHPGIGSSNHTDDSPYSWSFPPYPHPSSSEKYENTKTCRGPDELHA 180  
 Db 181 NLLCPVDVLDVPEGTLPDKQSTEQAIQLLEKMKTSASPFLLAVGYHKPHIPRYPKFQK 240  
 QY 181 NLLCPVDVLDVPEGTLPDKQSTEQAIQLLEKMKTSASPFLLAVGYHKPHIPRYPKFQK 240  
 Db 241 LYPLENITLAPDPEVPGDGLPPVAYNPWMDIRQEDVQALNLSVPGPLPVDFQKRIQSY 300  
 QY 241 LYPLENITLAPDPEVPGDGLPPVAYNPWMDIRQEDVQALNLSVPGPLPVDFQKRIQSY 300  
 Db 301 FASVSYLDTQVGRLLSALDDLOLANSTIIAFTSDHGALGHEGWAKYSNFDVATHVPLI 360  
 QY 301 FASVSYLDTQVGRLLSALDDLOLANSTIIAFTSDHGALGHEGWAKYSNFDVATHVPLI 360  
 Db 361 FYVPGRTASLPEAGEKLPFYLDPPDSASQLMEPGRQSDMLVELVSLPPTLAGLAGLOVPP 420  
 QY 361 FYVPGRTASLPEAGEKLPFYLDPPDSASQLMEPGRQSDMLVELVSLPPTLAGLAGLOVPP 420  
 Db 421 RCPVPSFHVLCRGGKLLKHFRFRDLEEDPYLPGNPRELIAYSQYPRPSDIPQWNSDKP 480  
 QY 421 RCPVPSFHVLCRGGKLLKHFRFRDLEEDPYLPGNPRELIAYSQYPRPSDIPQWNSDKP 480  
 Db 481 SLKDIKIMGYSIRTIIDYRYTVWVGFNDFEFLANFSDIHAGELYFVDSPLQDHNMYNDSQ 540  
 QY 481 SLKDIKIMGYSIRTIIDYRYTVWVGFNDFEFLANFSDIHAGELYFVDSPLQDHNMYNDSQ 540  
 Db 541 GGDLFQLLMP 550  
 QY 541 GGDLFQLLMP 550

## RESULT 3

ID Y02687 standard; Protein; 406 AA.  
 AC Y02687;  
 DT 11-JUN-1999 (first entry)  
 DE Human secreted protein encoded by gene 38 clone HSLB69.  
 KW Human; secreted protein; fusion protein; gene therapy; protein therapy;  
 KW diagnosis; tissue; cancer; tumour; neurodegenerative disorder; leukaemia;  
 KW developmental abnormality; foetal deficiency; blood; allergy; renal;  
 KW immune system; asthma; lymphocytic disease; brain; hepatic; lymphoma;  
 KW inflammation; ischaemic shock; Alzheimer's disease; restenosis; AIDS;  
 KW cognitive disorder; schizophrenia; prostate; obesity; osteoclast; thymus;  
 KW osteoporosis; arthritis; testis; lung; thyroiditis; thyroid; digestion;  
 KW endocrine; metabolism; regulation; malabsorption; gastritis; neoplasm.  
 OS Homo sapiens.  
 PN WO9902546-A1.  
 PD 21-JAN-1999.  
 PF 07-JUL-1998; U13684.  
 PR 12-SEP-1997; US-058785.  
 PR 08-JUL-1997; US-051916.  
 PR 08-JUL-1997; US-051918.  
 PR 08-JUL-1997; US-051919.  
 PR 08-JUL-1997; US-051920.  
 PR 08-JUL-1997; US-051925.







CC terminal peptide used in an example from the present invention. The  
 CC recombinant IDS is used to treat IDS deficiency where heparin sulphate  
 CC and dermatan sulphate accumulates in lysosomes resulting in Hunter  
 CC syndrome which is manifested by e.g. severe mental retardation,  
 CC skeletal deformities and stiff joints. The recombinant IDS possesses  
 CC inter alia an improved half life inter alia and improved uptake  
 CC properties in comparison to the naturally glycosylated molecule.  
 SQ Sequence 25 AA;

Query Match 3.3%; Score 136; DB 1; Length 25;  
 Best Local Similarity 84.0%; Pred. No. 1.05e+01;  
 Matches 21; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

Db 1 TSALNVLIIIVDDLRPSIGDYDDVL 25  
 QY 34 TDALNVLIIIVDDLRPSIGDYDDVL 58

## RESULT 9

ID W41995 standard; Protein; 226 AA.  
 AC W41995;  
 DT 02-JUL-1998 (first entry)  
 DE Flea cysteine protease SEQ ID NO:8.  
 KW Flea; Siphonaptera; protease; vaccine; prophylaxis; infestation;  
 KW Immunoglobulin protease; larvae; host animal.  
 OS Siphonaptera.  
 PN WO9740058-A1.  
 PD 30-OCT-1997.  
 PF 24-APR-1997; U06121.  
 PR 04-APR-1997; US-042945.  
 PR 24-APR-1996; US-639075.  
 PR 15-NOV-1996; US-749699.  
 PA (HESK-) HESKA CORP.  
 PI Frank GR, Gaines PJ, Grieve RB, Hunter SW, Rushlow KE,  
 PI Silver G, Stiegler GL;  
 DR WPI: 98-076762/07.  
 DR N-PSDB: V04624.  
 PT New flea protease genes and proteins - used in vaccine compositions  
 PT for the prophylaxis and treatment of flea infestation, especially in  
 PT cats or dogs  
 PS Claim 2; Page 174-175; 318pp; English.  
 CC The present sequence represents a novel flea cysteine protease. The  
 CC protease, its mimetopes, antibodies (Ab) and inhibitors of the  
 CC protein, as well as the DNA encoding the protein, may all be used in  
 CC therapeutic compositions to reduce flea protease activity (especially  
 CC immunoglobulin protease) and so reduce flea infestation, especially in  
 CC cats or dogs. Alternatively, flea larvae may ingest the faeces of adult  
 CC fleas which comprises anti-protease antibodies produced by a host animal  
 CC in response to administration of the protein. Therapeutic compositions  
 CC may further comprise a compound that reduces haematophagous ectoparasite  
 CC burden by a method other than by reducing flea immunoglobulin protease  
 CC activity. The novel flea DNA encoding the protein can also be used to  
 CC produce recombinant protein, and fragments of it are used as probes and  
 CC primers for identification and isolation of related sequences, also as  
 CC antisense, triplex-forming agents and ribozymes for inhibition of the  
 CC synthesis of the protein. Ab are also useful for screening expression  
 CC libraries, to purify the protein and to target cytotoxins to fleas.  
 SQ Sequence 226 AA;

Query Match 2.7%; Score 109; DB 1; Length 226;  
 Best Local Similarity 25.3%; Pred. No. 9.52e+00;  
 Matches 22; Conservative 26; Mismatches 35; Indels 4; Gaps 4;

Db 119 RNDVDMKVALKHGPISTVAIDASHKTSFYSNGVYQPKGNKRGOLDHVLVVG-YGE 177  
 QY 273 REDVQALNISV-PYGPVPDFQKIRQS-YFASVSYLDTQVGRLLSALDQLANSTIIA 330

Db 178 INSEPYWLKPNQGLMGNVIFDVAKN 204  
 QY 331 FTSDHGVALGEGE-WAKYSNFDVATH 356

## RESULT 10

ID W41994 standard; Protein; 327 AA.  
 AC W41994;  
 DT 02-JUL-1998 (first entry)  
 DE Flea cysteine protease SEQ ID NO:5.  
 KW Flea; Siphonaptera; protease; vaccine; prophylaxis; infestation;  
 KW Immunoglobulin protease; larvae; host animal.  
 OS Siphonaptera.  
 PN WO9740058-A1.  
 PD 30-OCT-1997.  
 PF 24-APR-1997; U06121.  
 PR 04-APR-1997; US-042945.  
 PR 24-APR-1996; US-639075.  
 PR 15-NOV-1996; US-749699.  
 PA (HESK-) HESKA CORP.  
 PI Frank GR, Gaines PJ, Grieve RB, Hunter SW, Rushlow KE,  
 PI Silver G, Stiegler GL;  
 DR WPI: 98-076762/07.  
 DR N-PSDB: V04620.  
 PT New flea protease genes and proteins - used in vaccine compositions

Query Match 2.7%; Score 109; DB 1; Length 327;  
 Best Local Similarity 25.3%; Pred. No. 9.52e+00;  
 Matches 22; Conservative 26; Mismatches 35; Indels 4; Gaps 4;

Db 220 RNDVDMKVALKHGPISTVAIDASHKTSFYSNGVYQPKGNKRGOLDHVLVVG-YGE 278  
 QY 273 REDVQALNISV-PYGPVPDFQKIRQS-YFASVSYLDTQVGRLLSALDQLANSTIIA 330

Db 279 INSEPYWLKPNQGLMGNVIFDVAKN 305  
 QY 331 FTSDHGVALGEGE-WAKYSNFDVATH 356

## RESULT 11

ID W41993 standard; Protein; 327 AA.  
 AC W41993;  
 DT 02-JUL-1998 (first entry)  
 DE Flea cysteine protease SEQ ID NO:2.  
 KW Flea; Siphonaptera; protease; vaccine; prophylaxis; infestation;  
 KW Immunoglobulin protease; larvae; host animal.  
 OS Siphonaptera.  
 PN WO9740058-A1.  
 PD 30-OCT-1997.  
 PF 24-APR-1997; U06121.  
 PR 04-APR-1997; US-042945.  
 PR 24-APR-1996; US-639075.  
 PR 15-NOV-1996; US-749699.  
 PA (HESK-) HESKA CORP.  
 PI Frank GR, Gaines PJ, Grieve RB, Hunter SW, Rushlow KE,  
 PI Silver G, Stiegler GL;  
 DR WPI: 98-076762/07.  
 DR N-PSDB: V04620.

Db 279 INSEPYWLKPNQGLMGNVIFDVAKN 305  
 QY 331 FTSDHGVALGEGE-WAKYSNFDVATH 356

Db 279 INSEPYWLKPNQGLMGNVIFDVAKN 305  
 QY 331 FTSDHGVALGEGE-WAKYSNFDVATH 356



	Query Match	2.6%	Score 107;	DB 1;	Length 4536;
	Best Local Similarity	28.0%;	Pred. No. 1.31e+01;		
	Matches	23;	Conservative	20;	Mismatches 34; Indels 5; Gaps 5;
D <b>b</b>	1265	NSLXI-EIPLPFGGKSSRDLMLEVTTPALHF-KSVGFHLPSREQVPTFTIPKLYQLQ	1322		
	:	:	:	:   :    :   :	:
Q <b>y</b>	187	DVLVDVEGTLPDKSQTEQAQLLEKMKTASPFELAVGYHKHPFRYPK-EFQKLYPLE	245		
	:	:	:	:   :     :	:
D <b>b</b>	1323	-VPILGLVDLDLSTNVSNLYN-W	1342		
	:	:	:	:   :    :	:
Q <b>y</b>	246	NITLAPDFEVPDGLPPVAYNPW	267		
	:	:	:	:   :    :	:

\*\*\*\*\*

W P S R E H (TM)

\*\*\*\*\*

Release 3.1A John F. Collins, Biocomputing Research Unit.  
Copyright (c) 1993-1998 University of Edinburgh, U.K.  
Distribution rights by Oxford Molecular Ltd

MPSrch\_pp protein - protein database search, using Smith-Waterman algorithm

Run on: Wed Apr 19 21:42:37 2000; MasPar time 8.07 Seconds  
Tabular output not generated. 137.498 Million cell updates/sec

Title: >US-09-249-003-3  
Description: (1-16) from US09249003.pep  
Perfect Score: 117  
Sequence: 1 PRELIATSNVPRNIP 16

Scoring table: PAM 150  
Gap 15

Searched: 225878 seqs, 69334122 residues

Post-processing: Minimum Match 0%  
Listing first 45 summaries

Database: sptrmb112  
1:sp\_archaea 2:sp\_bacteria 3:sp\_fungi 4:sp\_human  
5:sp\_invertebrate 6:sp\_mammal 7:sp\_mhc 8:sp\_organelle  
9:sp\_phage 10:sp\_plant 11:sp\_rodent 12:sp\_unclassified  
13:sp\_vertebrate 14:sp\_virus

Statistics: Mean 27.130; Variance 37.308; scale 0.727

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match %	ID	Description	Pred. No.
1	64	54.7	332	UVR REPAIR PROTEIN.	7.18e-01
2	64	54.7	416	IMPB/MUCB/SAMB FAMILY	7.18e-01
3	64	54.7	416	UV-DAMAGE REPAIR PROTE	7.18e-01
4	63	53.8	192	ESPA.	1.12e+00
5	63	53.8	192	ESPA PROTEIN.	1.12e+00
6	63	53.8	192	SECRETED PROTEIN ESPA.	1.12e+00
7	63	53.8	787	IRON(III) DICITRATE TR	1.12e+00
8	63	53.8	792	IRON(III) DICITRATE TR	1.12e+00
9	62	53.0	434	FLAGELLUM-SPECIFIC ATP	1.73e+00
10	62	53.0	434	ATPASE.	1.73e+00
11	62	53.0	454	MITOCHONDRIAL ELONGATI	1.73e+00
12	62	53.0	471	MITOCHONDRIAL ELONGATI	1.73e+00
13	60	51.3	192	ESPA.	4.10e+00
14	60	51.3	212	MITOCHONDRIAL ELONGATI	4.10e+00
15	60	51.3	259	ORF1.	4.10e+00
16	60	51.3	265	HYPOTHEICAL 31.0 KD P	4.10e+00
17	60	51.3	370	PRE-MRNA SPLICING PROT	4.10e+00
18	60	51.3	755	EXTRACELLULAR MATRIX P	4.10e+00
19	59	50.4	192	ESPA.	6.26e+00
20	59	50.4	1215	DNA FOR SEROTYPE B CAP	6.26e+00

21	58	49.6	160	1	Q9YAF4	160AA LONG HYPOTHETICA	9.51e+00
22	57	48.7	598	3	Q05930	CHROMOSOME XII COSMID	9.51e+00
23	57	48.7	90	10	Q9ZRN9	LECTIN (FRAGMENT).	1.44e+01
24	57	48.7	115	2	O31841	YOEK PROTEIN.	1.44e+01
25	57	48.7	256	2	O34760	YTNP.	1.44e+01
26	57	48.7	331	2	Q9ZCB6	HYPOTHETICAL 36.5 KD P	1.44e+01
27	57	48.7	624	3	O12246	ORF YOR171C.	1.44e+01
28	57	48.7	1271	2	O54504	MICROCYSITIN SYNTHETASE	1.44e+01
29	57	48.7	2166	14	O36637	POLYMERASE.	1.44e+01
30	57	48.7	2166	14	O36635	POLYMERASE.	1.44e+01
31	56	47.9	99	14	O81975	PAPILLOMAVIRUS (5A3) E	2.17e+01
32	56	47.9	108	1	O58630	108AA LONG HYPOTHETICA	2.17e+01
33	56	47.9	157	2	O85405	HYPOTHETICAL 18.0 KD P	2.17e+01
34	56	47.9	186	14	Q9YVJ7	ORF MSV245 PUTATIVE RN	2.17e+01
35	56	47.9	299	13	O92153	HAT-2.	2.17e+01
36	56	47.9	401	13	O93408	HOMEOBOX PROTEIN BIX1.	2.17e+01
37	56	47.9	401	13	O9YHW3	VACUOLAR-TYPE H+-ATPAS	2.17e+01
38	56	47.9	619	5	O96065	SEMAPHORIN H PRECURSOR	2.17e+01
39	56	47.9	775	11	P70275	SEMAPHORIN H PRECURSOR	2.17e+01
40	56	47.9	775	4	O15041	COLLAPSIDIN 5 PRECURSOR	2.17e+01
41	56	47.9	785	13	O42237	CLASS A CALCIUM CHANNE	2.17e+01
42	56	47.9	791	11	O70368	KIAA1007 PROTEIN (FRAG	2.17e+01
43	56	47.9	1089	4	O9Y2L0	ENAMELIN PRECURSOR.	2.17e+01
44	56	47.9	1142	6	O97939	SIMILAR TO HUMAN SREBP	2.17e+01
45	56	47.9	1758	5	Q22830		

ALIGNMENTS

RESULT 1  
ID O30670 PRELIMINARY; PRT; 332 AA.  
AC O30670;  
DT 01-JAN-1998 (TREMBLrel. 05, Created)  
DT 01-JAN-1998 (TREMBLrel. 05, Last sequence update)  
DT 01-NOV-1998 (TREMBLrel. 08, Last annotation update)  
DE UVR REPAIR PROTEIN.  
GN UVRX.  
OS Bacillus subtilis.  
OC Bacteria; Firmicutes; Bacillus/Clostridium group;  
OC Bacillus/Staphylococcus group; Bacillus.  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=168;  
RA PAIK S.H., HANSEN J.N.;  
RL Submitted (JUL-1997) to the EMBL/GenBank/DBJ databases.  
DR EMBL; AF014938; AAC63530.1;  
DR PFAM; PF00817; IMS; 1.  
SQ SEQUENCE 332 AA; 37768 MW; 7F4624B7 CRC32;

Query Match 54.7%; Score 64; DB 2; Length 332;  
Best Local Similarity 58.3%; Pred. No. 7.18e-01;  
Matches 7; Conservative 3; Mismatches 2; Indels 0; Gaps 0;

Db 1 MIDYSQPRKNI 12  
QY 4 LIATSNVPRNII 15

RESULT 2  
ID O64031 PRELIMINARY; PRT; 416 AA.  
AC O64031;  
DT 01-AUG-1998 (TREMBLrel. 07, Created)  
DT 01-AUG-1998 (TREMBLrel. 07, Last sequence update)  
DE IMPB/MUCB/SAMB FAMILY PROTEIN.  
GN YOLE.  
OS Bacteriophage SPBC2.  
OC Viruses; dsDNA viruses, no RNA stage; Tailed phages; Siphoviridae.  
RN [1]  
RP SEQUENCE FROM N.A.  
RA LAZAREVIC V., DUESTERHOEF A., SOLDO B., HILBERT H., MAUEL C.,  
RA KARAMATA D.;  
RL Submitted (AUG-1997) to the EMBL/GenBank/DBJ databases.

DR EMBL; AF020713; AAC12990.1; -;  
 DR PFAM; PF00817; IMS; 1.11e-01;  
 SQ SEQUENCE 416 AA; 46729 MW; 4856E2CF CRC32;

Query Match 54.7%; Score 64; DB 9; Length 416;  
 Best Local Similarity 58.3%; Pred. No. 7.18e-01;  
 Matches 7; Conservative 3; Mismatches 2; Indels 0; Gaps 0;

Db 1 MIDYQPPRKN1 12

:| | | | | | | |

QY 4 LIAYSNYPRNNI 15

RESULT 3  
 ID O31990 PRELIMINARY; PRT; 416 AA.  
 AC O31990;  
 DT 01-JAN-1998 (TREMBlrel. 05, Created)  
 DT 01-JAN-1998 (TREMBlrel. 05, Last sequence update)  
 DT 01-NOV-1998 (TREMBlrel. 08, Last annotation update)  
 DE UV-DAMAGE REPAIR PROTEIN.  
 GN UVRX.  
 OS Bacillus subtilis.  
 OC Bacteria; Firmicutes; Bacillus/Clostridium group;  
 OC Bacillus/Staphylococcus group; Bacillus.  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=168;  
 RX MEDLINE; 98044033.

RA KUNST F., OGASAWARA N., MOSZER I., ALBERTINI A.M., ALLONI G., AZEVEDO V., BERTERO M.G., BESSIERES P., BOLOTIN A., BORCHERT S., BORRIS R., BOURST L., BRANS A., BRAUN M., BRIGNELL S.C., BRON S., BROUILLET S., BRUSCHI C.V., CALDWELL B., CAPUANO V., CARTER N.M., CHOI S.K., CODANI J.J., CONNERTON I.F., CUMMINGS N.J., DANTEL R.A., DENIZOT F., DEVINE K.M., DUSTERHOFT A., EHRICH S.D., EMERSON P.T., ENTIAN K.D., ERRINGTON J., FABRET C., FERRARI E., FOULGER D., FRITZ C., FUJITA M., FUJITA Y., FUMA S., GALIZZI A., GALLERON N., GHIM S.Y., GLASER P., GOFFEAU A., GOLIGHTY E.J., GRANDI G., GUISEPPI G., GUY B.J., HAGA K., HAIECH J., HARWOOD C.R., HENAUT A., HILBERT H., HOLSAPPEL S., HOSONO S., HULLO M.F., ITAYA M., JONES L., JORIS B., KARAMATA D., KASAHARA Y., KLAER-BLANCHARD M., KLEIN C., KOBAYASHI Y., KOETTER P., KONINGSSTEIN G., KROGH S., KUMANO M., KURIITA K., LAPIDUS A., LARDINOIS S., LAUBER J., LAZAREVIC V., LEE S.M., LEVINE A., LIU H., MASUDA S., MAUEL C., MEDIGUE C., MEDINA N., MELLADO R.P., MIZUNO M., MOESTL D., NAKAI S., NOBACK M., NOONE D., O'REILLY M., OGAWA K., OGIMURA A., OUDEGA B., PARK S.H., PARRO V., POHL T.M., PORTELLE D., PORWOLLIK S., PRESCOTT A.M., PRESECAN E., PUIC P., PURNELLE B., RAPOPORT G., REY M., REYNOLDS S., RIEGER M., RIVOLTA C., ROCHA E., ROCHE B., ROSE M., SADAIE Y., SATO T., SCANLAN E., SCHLEICH S., SCHROETER R., SCOFFONE F., SEKIGUCHI J., SEKOWSKA A., SERO S.J., SERROR P., SHIN B.S., SOLDI B., SOROKIN A., TACCONI E., TAKAGI T., TAKAHASHI H., TAKEMARU K., TAKEUCHI M., TAKAKOSHI A., TANAKA T., TERPSTRA P., TOGNONI A., TOSATO V., UCHIYAMA S., VANDENBOL M., VANNIER F., VASSAROTTI A., VIARI A., WAMBUIT R., WEDLER E., WEDLER H., WEITENEGGER T., WINTERS P., WIPAT A., YAMAMOTO H., YAMANE K., YASUMOTO K., YATA K., YOSHIDA K., YOSHIKAWA H.F., ZUMSTEIN E., YOSHIKAWA H., DANCHIN A.;  
 RT "The complete genome sequence of the gram-positive bacterium Bacillus subtilis.";  
 RL Nature 390:249-256(1997).  
 RN [2]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=168;  
 RA KUNST F., OGASAWARA N., YOSHIKAWA H., DANCHIN A.;  
 RL Submitted (NOV-1997) to the EMBL/GenBank/DBJ databases.  
 DR EMBL; Z99115; CAB14068.1; -;  
 DR PFAM; PF00817; IMS; 1.  
 SQ SEQUENCE 416 AA; 46729 MW; 4856E2CF CRC32;

Query Match 54.7%; Score 64; DB 2; Length 416;  
 Best Local Similarity 58.3%; Pred. No. 7.18e-01;  
 Matches 7; Conservative 3; Mismatches 2; Indels 0; Gaps 0;

Db 1 MIDYQPPRKN1 12

:| | | | | | | |

QY 4 LIAYSNYPRNNI 15

QY 4 LIAYSNYPRNNI 15

:| | | | | | | |

RESULT 4  
 ID O87743 PRELIMINARY; PRT; 192 AA.  
 AC O87743;  
 DT 01-NOV-1998 (TREMBlrel. 08, Created)  
 DT 01-NOV-1998 (TREMBlrel. 08, Last sequence update)  
 DT 01-NOV-1998 (TREMBlrel. 08, Last annotation update)  
 DE ESPA.  
 GN ESPA.  
 OS Escherichia coli.  
 OC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;  
 OC Escherichia.  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=19;  
 RA FRANKEL G.;  
 RL Submitted (MAR-1998) to the EMBL/GenBank/DBJ databases.  
 RN [2]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=19;  
 RA NEVES B.N.;  
 RT "Molecular and ultrastructural characterization of EspA from different enteropathogenic Escherichia coli serotypes.";  
 RL Submitted (AUG-1998) to the EMBL/GenBank/DBJ databases.  
 DR EMBL; AJ225018; CAA12348.1; -;  
 SQ SEQUENCE 192 AA; 20533 MW; 6FAE87F1 CRC32;

Query Match 53.8%; Score 63; DB 2; Length 192;  
 Best Local Similarity 60.0%; Pred. No. 1.12e-00;  
 Matches 9; Conservative 3; Mismatches 3; Indels 0; Gaps 0;

Db 104 QEVVDYINDPRNDI 118

:| | | | | | | |

QY 1 PRELIAYSNYPRNNI 15

RESULT 5  
 ID O69412 PRELIMINARY; PRT; 192 AA.  
 AC O69412;  
 DT 01-AUG-1998 (TREMBlrel. 07, Created)  
 DT 01-AUG-1998 (TREMBlrel. 07, Last sequence update)  
 DT 01-NOV-1998 (TREMBlrel. 08, Last annotation update)  
 DE ESPA PROTEIN.  
 GN ESPA.  
 OS Escherichia coli.  
 OC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;  
 OC Escherichia.  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN-EHC EDL933;  
 RA KRISSE A.U., EBEL F., DEIBEL C., CHAKRABORTY T., GUZMAN C.A.;  
 RL Submitted (MAY-1997) to the EMBL/GenBank/DBJ databases.  
 RN [2]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN-EDL933;  
 RX MEDLINE; 98339885.

RA PERNA N.T., MAYHEW G.F., POSFAI G., ELLIOTT S., DONNENBERG M.S., KAPER J.B., BLATTNER F.R.;  
 RT "Molecular evolution of a pathogenicity island from enterohemorrhagic Escherichia coli O157:H7.";  
 RL Infect. Immun. 66:3810-3817(1998).  
 DR EMBL; Y13068; CAA73506.1; -;  
 DR EMBL; AF071034; AAC31501.1; -;  
 SQ SEQUENCE 192 AA; 20574 MW; 65635B9E CRC32;

Query Match 53.8%; Score 63; DB 2; Length 192;  
 Best Local Similarity 50.0%; Pred. No. 1.12e-00;  
 Matches 8; Conservative 5; Mismatches 3; Indels 0; Gaps 0;

Db 104 QDVVDYINDPRNDIS 119



```

DT 01-MAY-1999 (Tremblrel. 10, Last annotation update)
DE FLAGELLUM-SPECIFIC ATP SYNTHASE.
GN FLII.
OS Helicobacter pylori J99.
OC Bacteria; Proteobacteria; epsilon subdivision; Helicobacter group;
OC Helicobacter.
[1]
RP SEQUENCE FROM N.A.
RC STRAIN-J99;
RX MEDLINE; 99120557.
RA ALM R.A., LING L.-S.L., MOIR D.T., KING B.L., BROWN E.D., DOIG P.C.,
RA SMITH D.R., NOONAN B., GUILD B.C., DEJONGE B.L., CARMEL G.,
RA TOMMINS P.J., CARUSO A., ORIA-NICKELSEN M., MILLS D.M., IVES C.,
RA GIBSON R., MERBERG D., MILLS S.D., JIANG Q., TAYLOR D.E., VOVIS G.F.,
RA TRUST T.J.;
RT "Genomic-sequence comparison of two unrelated isolates of the human
RL Nature 397:176-180(1999).
DR EMBL; AE001554; AAD06888.1; -.
SQ SEQUENCE 434 AA; 47703 MW; A4523E49 CRC32;

Query Match 53.0%; Score 62; DB 2; Length 434;
Best Local Similarity 53.3%; Pred. No. 1.73e+00;
Matches 8; Conservative 2; Mismatches 5; Indels 0; Gaps 0;

Db 334 SRELDYGIYPPINI 348
QY 1 PRELIAYSNYPNNI 15
:|||||:|:|:|

RESULT 10
ID 054374 PRELIMINARY; PRT; 434 AA.
AC 054374;
DT 01-JUN-1998 (Tremblrel. 06, Created)
DT 01-JUN-1998 (Tremblrel. 06, Last sequence update)
DT 01-MAY-1999 (Tremblrel. 10, Last annotation update)
DE ATPASE.
GN FLII.
OS Helicobacter pylori (Campylobacter pylori).
OC Bacteria; Proteobacteria; epsilon subdivision; Helicobacter group;
OC Helicobacter.
[1]
RP SEQUENCE FROM N.A.
RC STRAIN-CCUG 17874;
RA PORWOLLIK S., O'TOOLE P.W.;
RL Submitted (OCT-1996) to the EMBL/GenBank/DBJ databases.
DR EMBL; U75584; AAB93498.1; -.
DR PFAM; PF00006; ATP-synt_ab; 1.
SQ SEQUENCE 434 AA; 47592 MW; 221EA346 CRC32;

Query Match 53.0%; Score 62; DB 2; Length 434;
Best Local Similarity 53.3%; Pred. No. 1.73e+00;
Matches 8; Conservative 2; Mismatches 5; Indels 0; Gaps 0;

Db 334 SRELDYGIYPPINI 348
QY 1 PRELIAYSNYPNNI 15
:|||||:|:|:|

RESULT 11
ID 092T91 PRELIMINARY; PRT; 454 AA.
AC 092T91;
DT 01-MAY-1999 (Tremblrel. 10, Created)
DT 01-MAY-1999 (Tremblrel. 10, Last sequence update)
DT 01-NOV-1999 (Tremblrel. 12, Last annotation update)
DE MITOCHONDRIAL ELONGATION FACTOR TU.
GN TUFA OR T419.19 OR T5J8.25.
OS Arabidopsis thaliana (Mouse-ear cress).
OC Eukaryota; Viridiplantae; Streptophyta; Tracheophyta;
OC euphyllophytes; Spermatophyta; Magnoliophyta; eudicotyledons;
OC core eudicots; Rosidae; eurosids II; Brassicales; Brassicaceae;
OC Arabidopsis.
[1]

```

```

RP SEQUENCE FROM N.A.
RC STRAIN-CV. COLUMBIA;
RA PARNEILL L.D., GNOJ L., DE LA BASTIDE M., HAMEED A., HABERMANN K.,
RA SCHUTZ K., HUANG E., GOTTESMAN T., DEDHIA N.N., MCCOMBIE W.R.;
RT "Genomic sequence of BAC T419 from Arabidopsis thaliana, Chromosome
RT IV, near 16.6 cm.";
RL Submitted (MAY-1998) to the EMBL/GenBank/DBJ databases.
[2]
RP SEQUENCE FROM N.A.
RC STRAIN-CV. COLUMBIA;
RA DE LA BASTIDE M., GNOJ L., HABERMANN K., HUANG E.N., GOTTESMAN T.,
RA KAPLAN N., LODHI M., JENSEN K., HAMEED A., SCHUTZ K., MARTIENSSEN R.,
RA DEDHIA N., PARNEILL L.D., MCCOMBIE W.R.;
RT "Arabidopsis thaliana BAC T5J8 from chromosome IV, short arm.";
RL Submitted (FEB-1999) to the EMBL/GenBank/DBJ databases.
DR EMBL; AF069442; AAC79113.1; -.
DR EMBL; AC004044; AAD15337.1; -.
DR HSSP; P02990; IEFU.
DR MENDEL; 34382; Arath;Tufa;34382.
DR PROSITE; PS00301; EFATOR_GTP; 1.
KW Elongation factor; Protein biosynthesis; GTP-binding.
SQ SEQUENCE 454 AA; 49410 MW; 3B0B22FE CRC32;

Query Match 53.0%; Score 62; DB 10; Length 454;
Best Local Similarity 40.0%; Pred. No. 1.73e+00;
Matches 6; Conservative 7; Mismatches 2; Indels 0; Gaps 0;

Db 210 RELISFYKPGDDIP 224
QY 2 RELIAYSNYPNNIP 16
|||||:|:|:|

RESULT 12
ID 039206 PRELIMINARY; PRT; 471 AA.
AC 039206;
DT 01-NOV-1996 (Tremblrel. 01, Created)
DT 01-NOV-1996 (Tremblrel. 01, Last sequence update)
DT 01-NOV-1999 (Tremblrel. 12, Last annotation update)
DE MITOCHONDRIAL ELONGATION FACTOR TU PRECURSOR.
GN TUFA.
OS Arabidopsis thaliana (Mouse-ear cress).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC euphyllophytes; Spermatophyta; Magnoliophyta; eudicotyledons;
OC core eudicots; Rosidae; eurosids II; Brassicales; Brassicaceae;
OC Arabidopsis.
[1]
RP SEQUENCE FROM N.A.
RC STRAIN-CV. COLUMBIA;
RX MEDLINE; 96145515.
RA KUHLMAN P., PALMER J.D.;
RT "Isolation, expression, and evolution of the gene encoding
RT mitochondrial elongation factor Tu in Arabidopsis thaliana.";
RL Plant Mol. Biol. 29:1057-1070(1995).
[2]
RP SEQUENCE OF 359-471 FROM N.A.
RC TISSUE-CELL SUSPENSION CULTURE OF ECOTYPE COLUMBIA;
RA COOKE R., LAUDIE M., RAYNAL M., DELSENY M.;
RL Submitted (JUL-1995) to the EMBL/GenBank/DBJ databases.
DR EMBL; X89227; CAA61511.1; -.
DR EMBL; F14375; CAA23078.1; -.
DR HSSP; P02990; IEFU.
DR MENDEL; 13559; Arath;Tufa;13559.
DR PROSITE; PS00301; EFATOR_GTP; 1.
DR PFAM; PF00009; GTP_EFTU; 1.
KW Transit peptide; Elongation factor; Protein biosynthesis;
KW GTP-binding.
FT TRANSIT 1 72 POTENTIAL.
FT CHAIN 73 471 MITOCHONDRIAL ELONGATION FACTOR TU.
SQ SEQUENCE 471 AA; 51384 MW; 849FCE59 CRC32;

Query Match 53.0%; Score 62; DB 10; Length 471;
Best Local Similarity 40.0%; Pred. No. 1.73e+00;
Matches 6; Conservative 7; Mismatches 2; Indels 0; Gaps 0;

```



```

Db 227 RELSFKYKPGDDIP 241
   |||::: ::||
QY 2 RELIAYSNYPRNNIP 16

RESULT 13
ID Q47184 PRELIMINARY; PRT; 192 AA.
AC Q47184;
DT 01-NOV-1996 (TREMBlrel. 01, Created)
DT 01-NOV-1996 (TREMBlrel. 01, Last sequence update)
DT 01-MAY-1999 (TREMBlrel. 10, Last annotation update)
DE ESPA.
OS Escherichia coli.
GN Escherichia coli.
OC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
OC Escherichia.
RN [1]
RC SEQUENCE FROM N.A.
RX STRAIN-ENTEROPATHOGENIC SEROTYPE O127:H6;
RX MEDLINE; 96310369.
RT KENNY B., LAI L., FINLAY B., DONNENBERG M.S.;
RT "EspA, a protein secreted by enteropathogenic Escherichia coli, is
RT required to induce signals in epithelial cells.";
RL Mol. Microbiol. 20:313-323(1996).
RN [2]
RP SEQUENCE FROM N.A.
RA ELLIOTT S.J., WAINWRIGHT L.A., MCDANIEL T.K., JARVIS K.G., DENG Y.K.,
RA LAI L.C., MCNAMARA B.P., DONNENBERG M.S., KAPER J.B.;
RL Mol. Microbiol. 0:0-0(1998).
RN [3]
RP SEQUENCE OF 1-191 FROM N.A.
RC STRAIN-30;
RA FRANKEL G.;
RL Submitted (MAR-1998) to the EMBL/GenBank/DBJ databases.
RN [4]
RP SEQUENCE FROM N.A.
RC STRAIN-30;
RX MEDLINE; 99068022.
RA NEVES B.N., KNUSTON S., TRABULSI L.R., SPERANDIO V., KAPER J.B.,
RA DOUGAN G., FRANKEL G.;
RT "Molecular and ultrastructural characterisation of EspA from different
RT enteropathogenic Escherichia coli serotypes.";
RL FEMS Microbiol. Lett. 169:73-80(1998).
DR EMBL; Z54352; CAA91163.1; -
DR EMBL; AF022236; AAC38394.1; -
DR EMBL; AJ225019; CAA12349.1; -
SQ SEQUENCE 192 AA; 20469 MW; 8D6D33B9 CRC32;

Query Match 51.3%; Score 60; DB 2; Length 192;
Best Local Similarity 60.0%; Pred. No. 4.10e+00;
Matches 9; Conservative 3; Mismatches 3; Indels 0; Gaps 0;

Db 104 PDEVISYINDPRNDI 118
   |::|::|::|
QY 1 PRELIAYSNYPRNNI 15

RESULT 14
ID Q9ZR06 PRELIMINARY; PRT; 212 AA.
AC Q9ZR06;
DT 01-MAY-1999 (TREMBlrel. 10, Created)
DT 01-MAY-1999 (TREMBlrel. 10, Last sequence update)
DT 01-NOV-1999 (TREMBlrel. 12, Last annotation update)
DE MITOCHONDRIAL ELONGATION FACTOR TU (FRAGMENT).
GN TUFA.
OS Catharanthus roseus (Rosy periwinkle) (Madagascar periwinkle).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC euphyllophytes; Spermatophyta; Magnoliophyta; eudicotyledons;
OC core eudicots; Asteridae; euasterids I; Gentianales; Apocynaceae;
OC Catharanthus.
RN [1]
RP SEQUENCE FROM N.A.

```

```

RA SCHROEDER G., LURZ G., SCHROEDER J.;
RT "Mitochondrial elongation factor Tu from Catharanthus roseus.";
RL Submitted (NOV-1998) to the EMBL/GenBank/DBJ databases.
DR EMBL; AJ130966; CAA10267.1; -
DR HSP; P02990; LEFU.
DR MENDEL; 34753; Catro:Tufa;34753.
KW Elongation factor.
FT NON_TER 1
FT NON_TER 212
SQ SEQUENCE 212 AA; 23787 MW; 69A1E059 CRC32;

Query Match 51.3%; Score 60; DB 10; Length 212;
Best Local Similarity 40.0%; Pred. No. 4.10e+00;
Matches 6; Conservative 7; Mismatches 2; Indels 0; Gaps 0;

Db 11 RELSFKYKPGDEIP 25
   |||::: ::||
QY 2 RELIAYSNYPRNNIP 16

RESULT 15
ID Q44477 PRELIMINARY; PRT; 259 AA.
AC Q44477;
DT 01-NOV-1996 (TREMBlrel. 01, Created)
DT 01-NOV-1996 (TREMBlrel. 01, Last sequence update)
DT 01-NOV-1998 (TREMBlrel. 08, Last annotation update)
DE ORF1.
OS Azotobacter vinelandii.
OC Bacteria; Proteobacteria; gamma subdivision; Azotobacteraceae;
OC Azotobacter.
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-CA;
RA MYLONA P.V., RICKE S.C., BISHOP P.E.;
RL Submitted (OCT-1995) to the EMBL/GenBank/DBJ databases.
DR EMBL; U39675; AAB82617.1; -
DR PFAM; PF00005; ABC_tran; 1.
SQ SEQUENCE 259 AA; 28537 MW; C337B05C CRC32;

Query Match 51.3%; Score 60; DB 2; Length 259;
Best Local Similarity 41.7%; Pred. No. 4.10e+00;
Matches 5; Conservative 6; Mismatches 1; Indels 0; Gaps 0;

Db 96 RELVGFGRYPWH 107
   |||::: ||::
QY 2 RELIAYSNYPRN 13

Search completed: Wed Apr 19 21:44:12 2000
Job time : 95 secs.

```

\*\*\*\*\*

WIP5024 (TM)

\*\*\*\*\*

Release 2.1D John F. Collins, Biocomputing Research Unit.  
Copyright (c) 1993, 1994, 1995 University of Edinburgh, U.K.  
Distribution rights by IntelliGenetics, Inc.

MPsrch\_nn n.a. - n.a. database search, using Smith-Waterman algorithm

Run on: Wed Oct 4 13:48:44 1995; MasPar time 1618.89 Seconds  
1044.595 Million cell updates/sec

Tabular output not generated.

Title: >US-08-345-212-1  
Description: (1:2297) from US08345212.seq  
Perfect Score: 2297  
N.A. Sequence: 1 CGGCTGTGTGGAGTCTT.....TAATCATGTTTCTTTTTC 2297  
Comp: GCCCACACACCGCTCAGAA.....ATTAGGTACAAGAAAAAGG

Scoring table: TABLE default  
Gap 6

Nmatch STD : Dbase 0; Query 0

Searched: 493065 seqs, 368106350 bases x 2

Database: emb1-new6

- 1 BCT
- 2 EST1
- 3 EST2
- 4 EST3
- 5 EST4
- 6 EST5
- 7 FUN
- 8 INV1
- 9 INV2
- 10 MAM
- 11 ORG
- 12 PLN
- 13 PRI
- 14 PRO
- 15 ROD
- 16 STS
- 17 SYN
- 18 UNC
- 19 VRT
- 20 VIR

Database: EST

- 21 EST1
- 22 EST2
- 23 EST3
- 24 EST4
- 25 EST5

- 26 EST6
- 27 EST7
- 28 EST8
- 29 EST9
- 30 EST10
- 31 EST11
- 32 EST12
- 33 EST13
- 34 EST14
- 35 EST15
- 36 EST16
- 37 EST17
- 38 EST18
- 39 EST19
- 40 EST20
- 41 EST21
- 42 EST22
- 43 EST23
- 44 EST24
- 45 EST25
- 46 EST26
- 47 EST27
- 48 EST28
- 49 EST29
- 50 EST30
- 51 EST31
- 52 EST32
- 53 EST33
- 54 EST34
- 55 EST35
- 56 EST36
- 57 EST37
- 58 EST38
- 59 EST39
- 60 EST40
- 61 EST41
- 62 EST42
- 63 EST43
- 64 EST44
- 65 EST45
- 66 EST46
- 67 EST47
- 68 EST48
- 69 EST49
- 70 EST50
- 71 EST51
- 72 EST52
- 73 EST53
- 74 EST54
- 75 EST55
- 76 EST56
- 77 EST57
- 78 EST58

Database: genbank89

- 79 BCT1
- 80 BCT2
- 81 BCT3
- 82 BCT4
- 83 BCT5
- 84 BCT6
- 85 INV1
- 86 INV2
- 87 INV3

88 INV4  
89 INV5  
90 MAM1  
91 MAM2  
92 PAT1  
93 PAT2  
94 PHG  
95 PLN1  
96 PLN2  
97 PLN3  
98 PLN4  
99 PLN5  
100 PLN6  
101 PLN7  
102 PRI1  
103 PRI2  
104 PRI3  
105 PRI4  
106 PRI5  
107 PRI6  
108 PRI7  
109 PRI8  
110 PRI9  
111 ROD1  
112 ROD2  
113 ROD3  
114 ROD4  
115 ROD5  
116 ROD6  
117 ROD7  
118 STR  
119 STS1  
120 STS2  
121 STS3  
122 STS4  
123 SYN  
124 UNA  
125 VRL1  
126 VRL2  
127 VRL3  
128 VRL4  
129 VRL5  
130 VRL6  
131 VRT1  
132 VRT2  
133 VRT3

Database:

genbank-new6

134 BCT  
135 EST1  
136 EST2  
137 EST3  
138 EST4  
139 EST5  
140 EST6  
141 EST7  
142 EST8  
143 INV  
144 MAM  
145 PHG  
146 PLN  
147 PRI  
148 ROD  
149 STS

150 STR  
151 SYN  
152 UNA  
153 VRL  
154 VRT

Database: u-emb143\_89  
155 ALL

Statistics: Mean 12.082; Variance 3.168; scale 3.813

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description	Pred. No.
1	2297	100.0	2297	107 HUMIDSX	Human iduronate 2-sul	0.00e+00
2	1096	47.7	1750	113 MUSIDS	Mus musculus iduronat	0.00e+00
3	996	43.4	1120	107 HUMIDSGEN9	Homo sapiens iduronat	0.00e+00
4	995	43.3	36845	107 HUMIDS	Homo sapiens iduronat	0.00e+00
5	469	20.4	615	142 H06357	yl179h02.r1 Homo sapie	0.00e+00
6	469	20.4	615	3 HS357151	yl179h02.r1 Homo sapie	0.00e+00
7	434	18.9	1831	117 S64545	iduronate sulfatase I	0.00e+00
8	412	17.9	469	2 HS179158	yl185c04.r1 Homo sapie	0.00e+00
9	412	17.9	469	142 H05179	yl185c04.r1 Homo sapie	0.00e+00
c 10	369	16.1	378	56 T08577	EST06469 Homo sapiens	0.00e+00
c 11	360	15.7	484	38 R15374	yl179h02.r1 Homo sapie	0.00e+00
c 12	332	14.5	439	142 H06300	yl179h02.r1 Homo sapie	0.00e+00
c 13	332	14.5	439	3 HS300154	yl179h02.r1 Homo sapie	0.00e+00
c 14	321	14.0	327	26 HSC0SC011	H. sapiens partial cD	0.00e+00
c 15	319	13.9	423	71 T71926	ye07a10.r1 Homo sapie	0.00e+00
c 16	291	12.7	363	75 T87633	ye07a10.r1 Homo sapie	0.00e+00
c 17	290	12.6	297	31 HSC2PG011	ye07a10.r1 Homo sapie	0.00e+00
c 18	287	12.5	289	30 HSC2VA071	H. sapiens partial cD	0.00e+00
c 19	275	12.0	284	30 HSC2UC052	H. sapiens partial cD	0.00e+00
c 20	274	11.9	319	52 R64400	yl15h10.r1 Homo sapie	0.00e+00
c 21	258	11.2	261	30 HSC2UC051	H. sapiens partial cD	0.00e+00
c 22	258	11.2	310	46 R42115	yl15h10.r1 Homo sapie	0.00e+00
c 23	251	10.9	251	26 HSC0GA021	H. sapiens partial cD	0.00e+00
c 24	250	10.9	250	26 HSC0GA011	H. sapiens partial cD	0.00e+00
c 25	233	10.1	288	52 R64354	yl15a10.r1 Homo sapie	7.38e-304
c 26	231	10.1	233	24 HSAABBLT	H. sapiens putatively	1.03e-300
c 27	229	10.0	536	107 HUMIDSGEN1	Homo sapiens iduronat	1.42e-297
c 28	229	10.0	279	5 HS695141	yl185c09.r1 Homo sapie	1.42e-297
c 29	229	10.0	279	139 R79695	yl185c09.r1 Homo sapie	1.42e-297
c 30	221	9.6	237	30 HSC2VA072	H. sapiens partial cD	5.09e-285
c 31	218	9.5	238	26 HSC0SC012	H. sapiens partial cD	2.55e-280
c 32	214	9.3	282	52 R64401	yl15h10.r1 Homo sapie	4.67e-274
c 33	214	9.3	344	52 R64355	yl15a10.r1 Homo sapie	4.67e-274
c 34	205	8.9	322	108 HUMISS05	Homo sapiens iduronat	5.41e-260
c 35	205	8.9	490	107 HUMIDSGEN5	Homo sapiens iduronat	5.41e-260
c 36	186	8.1	202	65 T49494	ya75h11.r1 Homo sapie	1.98e-230
c 37	184	8.0	196	26 HSC0GA022	H. sapiens partial cD	2.49e-227
c 38	182	7.9	388	107 HUMIDSGEN3	Homo sapiens iduronat	3.13e-224
c 39	182	7.9	750	147 HUMIDS02	Homo sapiens iduronat	3.13e-224
c 40	182	7.9	750	13 HSIDS02	Homo sapiens iduronat	3.13e-224
c 41	182	7.9	298	108 HUMISS03	Homo sapiens iduronat	3.13e-224
c 42	178	7.7	290	108 HUMISS06	Homo sapiens iduronat	4.84e-218
c 43	178	7.7	379	107 HUMIDS06	Homo sapiens iduronat	4.84e-218
c 44	176	7.7	404	107 HUMIDSGEN8	Homo sapiens iduronat	5.97e-215



```
QY 961 AACAATCAGTGTGCGGATGGTCGAATTCCTGTGGACTTCACGGCGAAATGCCGCAGAG 1020
Db 1021 ctacttgccctctgtgcataattgatgatacacacaggtcgccgcctcttgagtgcttggga 1080
QY 1021 CTACTTTGCGCTCTGTGTCTATATTGGATACACAGGTGCGCGCGCTCTTGAGTGCCTTGA 1080
Db 1081 cgatcttcagctggccacagcaccatcattgcatctacotcggatcatgggtggctct 1140
QY 1081 CGATCTTCAGCTGGCCACAGCACCATCAITTCGATTTACCTCGGATCATGGTGGGCTCT 1140
Db 1141 agtgaactggagatggcccaaatcacagcaatttggatgttgcctaccatgttccct 1200
QY 1141 AGGTGAACATGAGAGATGGGCCAAATACAGCAATTTTGATGTGTGCTACCGATGTTCCCT 1200
Db 1201 gatctctatgttctcgaagagcugcttcacttcggagcgagcgagagcgttttccc 1260
QY 1201 GATATTCATGTCTCGAAGACGGCTTCACCTCCGAGCGAGCGAAGCTTTTCCC 1260
Db 1261 ttacctcgaccttttgattccgcctcacagttgatggagccagcgaggaatccatgga 1320
QY 1261 TTACCTCGACCTTTTGATTCGCCCTCACAGTTGATGGAGCGAGCGAGCAATCCATGGA 1320
Db 1321 ccttgtggaactgtgtctcttttcccacgtggtggactgcaggagtgaggttcc 1380
QY 1321 CCTTGTGGAACTTGTGTCTCTTTTCCACGCTGGCTGGACTTCGAGGACTGCAGGTTCC 1380
Db 1381 acctcgctgccccgttccttcttcatttcacgttgagctgtcgagagaaggcaagacctct 1440
QY 1381 ACCTCGCTGCCCGGTTCCTTCATTTTCAGCTGTGAGCTGTGACAGAGGCAAGAACCTTCT 1440
Db 1441 gaagcatttccagctcgtgacttgagagagatcgtacctccctcgttaatcccgctga 1500
QY 1441 GAAGCATTTTCGATTCGCTGACTTGAAGAGGATCCGTACCTCCCTGGTAAATCCCGCTGA 1500
Db 1501 actgatgcctatgcaagatccccggccttcagacatccctcagtggaattctgacaa 1560
QY 1501 ACTGATTCGCTATAGCCAGTATCCCGCGCTTCAGACATCCCTCAGTGAATTCGCACAA 1560
Db 1561 gccgagtttaaaagataaagatcatgggtattccatcacgaccatagactatagtta 1620
QY 1561 GCCGAGTTTAAAGATATAAAGATCATGGCTATTCATAGCCACCATAGACTATAGGTA 1620
Db 1621 tactgtgggttggttccatcctgatgaattctagctaaactttctgacatccatgc 1680
QY 1621 TACTGTGGGTGGCTTCGAATTCCTGATGAATTTCTAGCTAACTTTCTGACATCCATGC 1680
Db 1681 aggggaactgattttgtgattctgacctgaccatgcaggatcacatgatgaatgattc 1740
QY 1681 AGGGGAACCTGATTTTGTGGATTTCTGACCCATTCAGGAGTCAAAATATGTAATGATTC 1740
Db 1741 ccaaggtggagatctttccagcttgatgccttgagtttcccaaccatggatggcaa 1800
QY 1741 CCAAGGTGGAGATCTTTTCCAGTTGCTGATGCTTCAGTTCCTGAGTTCCTGAGTGGCNA 1800
Db 1801 atgtgatgtgctccctccagctggtgagaggaggttagagctggttctgtgatt 1860
QY 1801 ATGTGATGCTCCCTCCAGCTGGTGAGAGGAGGAGTTAGAGCTGGCTGCTTTTGATT 1860
Db 1861 acccataattggaagcagcttgagggctagtgtaataatccaaacatgcatacaaatggg 1920
QY 1861 ACCCATATATTTGGAGACGCTTGAGGCTAGTTAATCCAAACATGCATCAACAAATTTGG 1920
Db 1921 cctgagaatagtgaacagcaaaccttttcgtttagctttattataaaatttataatgg 1980
```

```
QY 1921 CTGGAATATGTACAGCCAAACCTTTTCTTAGTCTTTATTAATAATTATAATTGGT 1980
Db 1981 aattggaccagtttttttttaattccctctttttaaacaagttacggtctatttactg 2040
QY 1981 AATTGGACCACTTTTTTTAAATTTCCCTCTTTTAAACAGCTTACGGCTTATTTACTG 2040
Db 2041 ataataacaagcaaaccaactcaagttatgtcatcctttggatcacgaagaccataca 2100
QY 2041 AATAAATACAAAGCAACAACTCAAGTTATGTATACCTTTGGATACGAAGACCATACA 2100
Db 2101 taatacaacaataacattatcacacaagaatactttcattatttgggaatttagtgc 2160
QY 2101 TAATACCAACAATACATTAATACACAAGAATACCTTCATTATTTGTGGAAATTTAGTGC 2160
Db 2161 atttcaaaagtaatacatatcatacaactagccaccacactaaagtccctgattatttgg 2220
QY 2161 ATTCAAAAAGTAATCATATATCAAACTAGSCACCACACTAAGTTCCTGATTATTTGTT 2220
Db 2221 tataatttaataatatcttatgagccctatatattcaaaatattattgttaaacatgtaa 2280
QY 2221 TATAATTTAATAATATATCTTATGAGCCCTATATTCAAAATATTATGTTAAACAICTAA 2280
Db 2281 tcatgtttctttttcc 2297
QY 2281 TCATGTTTCTTTTCC 2297
RESULT 2
LOCUS MUSIDS 1750 bp mRNA ROD 19-AUG-1993
DEFINITION Mus musculus iduronate sulfatase (Ids) mRNA, complete cds.
ACCESSION L07921
KEYWORDS iduronate sulfatase.
SOURCE Mus musculus (library: lambda ZAP II of Stragene) female adult
thymus cDNA to mRNA.
ORGANISM Mus musculus
Eukaryota; Animalia; Chordata; Vertebrata; Mammalia; Theria;
Eutheria; Rodentia; Myomorpha; Muridae; Murinae.
REFERENCE 1 (bases 1 to 1750)
AUTHORS Daniele,A., Faust,C.J., Herman,G.E., Natale,P.D. and Ballabio,A.
TITLE Cloning and characterization of the cDNA for the murine iduronate
sulfatase gene
JOURNAL Genomics 16, 755-757 (1993)
MEDLINE 93315172
COMMENT NCBI gi: 349713
FEATURES location/Qualifiers
source 1..1750
/organism="Mus musculus"
/dev_stage="adult"
/sequenced_mol="cDNA to mRNA"
/sex="female"
/tissue_type="thymus"
/tissue_lib="lambda ZAP II of Stragene"
21..1712
/gene="Ids"
/product="Iduronate sulfatase"
CDS
/translation="MVVDVTLPGNHVQSHRTCARAIWRQLSFLILGFSFCALESAAQ
GNSATDALNILLIVDDLRPSLCYCKDKLVRSNPIDOLASHSLVFNQAFQAQVACAPS
RVSLTGRPDTRLIDFNSYWRHISGNFTIPQIFKENGIVTMVGKVFHPGIGSNH
SDDYPYSWFFPSPSEFYKTKCGDKLHANLPCVDVADVPEGLPKQSTPE
EATRLIEKWKTSGPSFFLAVGYHKPHIPFRYPKEFKLYPLENI TLAPDPHPDPLPP
VAYNPWMD IREDVQALNISVPYCP IPEDFQFKIRQSYFASVSYLDITQVGHVLSALD
```

DLRLAHNTIIAFTSDHGMALGHEWKNYSNFDWATRPMLLYVPRTAPLPAAQKUL  
FPYRDPEDPADMDADRHTEDLVELVSLFPTLAGIAGLPVLIGAPSLIFMISFAEKA  
RIFRISICSSWTKWSOTCLVPRBLJAYSOVPRPADPQWNSDKPTLNDIBSMDIYTHV  
DYRTVWVGDFDSEFLANSFDIHAGELYFLYSDP LQDHNVYNDSSHGELLHLSLRP\*

BASE COUNT  
ORIGIN

400 a 481 c 414 g 455 t

DB 113; Score 1096; Match 84.4%; QryMatch 47.7%; Pred. No. 0.00e+00;  
Matches 1382; Conservative 0; Mismatches 250; Indels 6; Gaps 5;

Db 109 tggcttaggctccttctgcatacgccctggagtcgcgagcgcaaggcaactctgctacag 168  
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

Qy 168 TGGTCTGACCTCCGCTGCGCTCGGATCCGAAACCGCAGGCCACTCGACCACAG 227  
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

Db 169 atgctctgaacatccttctgatcattgtggatgatctccgccctccctgggctgtatg 228  
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

Qy 228 ATGCTCTGAACGTTCTTCATCATCTGGATGACCTGCGGCCCTCCCTGGGCTGTATG 287  
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

Db 229 gagaagctagtgggtcccgcaacattgaccagctggcaatcccaatagcgtctcttc 288  
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

Qy 288 GGCATAAGCTGGTAGGTGCCCAATATTACCACTGGCATGCCACAGCCCTCCTCTCC 347  
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

Db 289 agaagctcttgctcagcaagcagtgctgcgccaccaagtgcagtgctccttcacccgga 348  
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

Qy 348 AGAATGCCCTTTCGGCAGCAGACAGTGTGCGCCCGCAGCGGGTTCTTTCTCATCGCA 407  
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

Db 349 ggagcctgcacaccaccgcgctgtatgacttcaattctactggaggtacattctggaa 408  
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

Qy 408 GAGACCTGACACACCGCGCTGTACGACTTCAACTTCCTACTGGA GGGTGCA CCGCTGGA 467  
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

Db 409 attttccaccatccccaactctcaagagaaatggctacgtgacatgtcggttggga 468  
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

Qy 468 ACTTCTCCACCATGCCCGAGTCTTCAAGGGAATGGCTATGTGACCATGTGCGGGGA 527  
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

Db 469 aagctcttcacccgtgtatctctccaatcacagtgatgattatccatatagctgtctt 528  
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

Qy 528 AAGTCTTTCACCGCTGGATATCTTCAACCATACCGATGATTCTCGGTATAGCTGCTCT 587  
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

Db 529 ttccacettatcatccatccctcagaaatatgaaacacctaagacgtgtaaggggcaag 588  
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

Qy 588 TTCCACCTTATCATCTCTCTCGAAGTATGAAACACTAAGACATGTCGAGGCCAG 647  
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

Db 589 atggaagctccacccaactctgttctgtcgtatgtgagatgtgcttgaaggtta 648  
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

Qy 648 ATGGGAACCTCCATCCCAACCTGCTTTGCCCTGTGGATGTCTGGATGTCCCGAGGGCA 707  
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

Db 649 ccttgctgacaacacagacacagaagaagccatcaggtgtgtggaagaagatgaacat 708  
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

Qy 708 CTTGCTGCAAAAACAGACACTGAGCAAGCCATACAGTGTCTGGAAAAGATGAAAAGCT 767  
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

Db 709 caggtagctccttctcctggcagtggtgggtaccacaagccgcata tcccttcagatacc 768  
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

Qy 768 CAGCCAGCTCCTTTCTTCGGCGGTGGGTATCATTAAGCCACACATGCCCTTCAGATACC 827  
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

Db 769 ccaaggaaattcagaagtgatctccttgaaacataaccctggctcgtgatcccatg 828  
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

Qy 828 CAAGGAATTCAGAGTGTATCTCTTGGAGAACATCACCTTGGCCCCCGATCCCGAGG 887  
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

Db 829 ttctctgatagcctaccacctagctagctacaaccctggatgatatacaggagagggaaag 888  
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

Qy 888 TCCCTGTAGCCCTACCCCTGTGGGCTACAAACCCCTTGATGAGCATCAGGCAACGGGAAG 947  
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

Db 889 atgtccaagccttaaacatcagtgctcctatggaccaattcctgaggaattttcacgga 948  
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

Qy 948 ACGTCCAAAGCCTTAAACATCAGTGTGCGGTATGGTCCAAATTCCTGTGAGACTTCACGGCA 1007  
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

Db 949 aaatccgcccagagctactttgcttctgtgctctatttgatactcaggtcgagacatgttt 1008  
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

Qy 1008 AAATCCGCCAGAGCTACTTTCCCTGTGTCTATATTTGGATACACAGGTGCGCGGCTCT 1067  
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

Db 1009 tgaagtgttggatgatcttctagcccaacacaaatattgtcttcacatctgatc 1068  
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

Qy 1068 TGAGTGTCTTTGGAGCATCTTCAGCTGGCCACAGCACCATTGCAATTTACCTCGGATC 1127  
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

Db 1069 acggatgggcccctagtgaaacgaggaatgggccaagtacagcaactttgatgttgcta 1128  
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

Qy 1128 ATGGTGGCGCTCTAGTGACATGAGAAATGGGCCAAATACAGCAATTTGATGTGCTA 1187  
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

Db 1129 cccgtgtgccctgatgctctatgtccctggaagcagcgcccaacttctcggcagggc 1188  
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

Qy 1188 CCATGTTCCTCCCTGATATCTATGTTCCTGGAAAGCAGGCTTCACTTCGGAGGCGAGCG 1247  
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

Db 1189 agaagctcttccataccgggagccctttgatccggcctcggatttgatggatgcaggca 1248  
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

Qy 1248 AGAAGCTTTTCCCTTACCTCGACCCCTTTTGATTCCGCTCACAGTTGATGCGACGCA 1307  
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

Db 1249 ggcacacaggagacctcgtagaactgtatctcttctcccaagctcgctggacttgcag 1308  
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

Qy 1308 GGCATCCATCGACCTTCTGGAACCTGTGTCTCTTTTCCACGCTGGCTGGACTTGCAG 1367  
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

Db 1309 gactgccagt--cctcctcgtgccccaccctcttcttctgcttgagctttgacagagaag 1367  
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

Qy 1368 GACTCGAGGTTCCACCTCGCTGCCCGTTCCTTCATTTACGTTGAGCTGCGAGAGAG 1427  
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

Db 1368 gccaaatcttcaagaagcatttgcagctcccatgacttggaaagagagccagacttgttg 1427  
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

Qy 1428 CCAAGAACCTTCTGAAGCATTTTTCGATTCCGTGACTTGGAAAGAGATCCGTACCTCCCTG 1487  
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

Db 1428 gta---ccccgggagtgatcgctatagcagtagccccggcctgcagatttccctcagt 1485  
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

Qy 1488 GTAATCCCGTGAACTGATTGCCCTATAGCCAGTATGCCCGGCTTCAGACATCCCTCAGT 1547  
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

Db 1486 ggaattctgacaagccaacgftaaatgataaagg--tcatgggat--ctatacgacg-- 1542  
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

Qy 1548 GGAATCTGCAAGCCGAGCTTTAAAGATATAAAGATCATGGGCTATTCCATAGGCA 1607  
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

Db 1543 tagactacaggtacactgtgtgggttggcttcgactcagtagtaatttctggctaactttt 1602  
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

Qy 1608 TAGACTATAGGTATACTGTGTGGGTGGCTTCAATCCTGATGAATTTCTAGCTAACTTTT 1667  
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

Db 1603 ctgatatccatgccggggaactctattttgcattctgaccccaactgcaggatcaaatg 1662  
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

Qy 1668 CTGACATCCATGCAGGGCACTGTATTTTGTGGATTCTGACCCATTGCGAGGATCAATA 1727  
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

Db 1663 tatataatgactcccaacatggaggtcttctccatcactgagggcctgagttctgcag 1722  
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

Qy 1728 TGCTAATGATTTCCCAAGTGGAGATCTTTTCCAGTTGTTGATGCCCTTGAGTTTGGCAA 1787  
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

Db 1723 cctcagtgggcacatag 1740  
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

Qy 1788 CAAATGATGCAAAATGTG 1805  
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

RESULT 3  
LOCUS HUMUSGN9 1120 bp DNA PRI 18-MAR-1994  
DEFINITION Homo sapiens iduronate-2-sulfatase (IDS) gene, complete cds.  
ACCESSION U13329  
KEYWORDS IDS gene; iduronate 2-sulfatase.

SEGMENT 9 of 9  
SOURCE Homo sapiens (human).  
ORGANISM Homo sapiens  
Eukaryota; Animalia; Chordata; Vertebrata; Mammalia; Theria;  
Eutheria; Primates; Haplorhini; Catarrhini; Hominoidea.  
REFERENCE 1 (bases 1 to 1120)  
AUTHORS Wilson, P.J., Meaney, C.A., Hopwood, J.J. and Morris, C.P.  
TITLE Sequence of the human iduronate 2-sulfatase (IDS) gene  
JOURNAL Genomics 17 (3), 773-775 (1993)  
MEDLINE 94063929  
COMMENT NCBI gi: 405203  
FEATURES  
    Location/Qualifiers  
        1..1120  
            /organism="Homo sapiens"  
            /sequenced\_mol="DNA"  
            /tissue\_lib="lambda Charon 35, ATCC No. LAOXN101"  
            Join(L13321:332..434, L13322:157..293, L13323:82..259,  
            L13324:131..219, L13325:125..325, L13326:88..258,  
            L13327:99..225, L13328:129..302, L128..600)  
            /feature="IDS"  
            /map="Xq27.3-q28"  
            /notes="NCBI gi: 405205"  
            /codon\_start=1  
            /product="iduronate-2-sulfatase"  
            /translacion="MPPRTGRGLMLGLVLSVVCVALGSETQANSTTDALNVLIIIV  
            DDLRLSCGCKLVSPNIDQLASHLFLQNAFAQAVCA SRVSELTGRDPTRL  
            YDFNSYRVRHAGNFSTIPQFKENGYSVMSGVKVFHFGISNHTDDSPYSWGFPPYHP  
            SSEKVTMTGRDGLMLANLLCPVDVLDVPEGTLPDKOSTEQAIOLLEKMKTSASP  
            FFLAVGCHKPHIPRYPKFKQKLYLENITLAPDFEVPDGLPPVAYNPWDINQEDV  
            QALN15VPYGP IPVDVFORK1ROSYSFVSVDVQVGRLLSALDDIQLANSTIIAFTSD  
            HGMALGEHGWAKTSNEFVATHVPLIFVYVGRVATSLPEAGEKILFYLDPEDSASQIME  
            PQRSMDVLVSLFPLTAGLAGLVPPPCVPSFHVLELCREKMLKHFRRFLDEED  
            PYLPNPRELIAYSQYPRSDIPQWNSDKPSLKD IKTMGYSIRTIDYRTYVWGFNPD  
            EFLANFSDIHAGELYFVDSDP LQDHNMYNDSQGDLEQLLMP"  
            order (L13328:303..404, 1..127)  
            /feature="IDS"  
            /map="Xq27.3-q28"  
            /number=8  
            128..>1120  
            /feature="IDS"  
            /map="Xq27.3-q28"  
            /number=9  
            /evidence=experimental  
BASE COUNT 318 a 232 c 212 g 358 t  
ORIGIN  
  
DB 107; Score 996; Match 100.0%; OryMatch 43.4%; Pred. No. 0.00e+00;  
Matches 996; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
  
Db 125 caggcaggcaatccatgacgtgtggaacttgtgtctttttccacgctggctgac 184  
|||||  
Qy 1302 CAGGCAGGCAATCCATGACCTGTGGAACCTGTGTCTCTTTCCACCGCTGGCTGCAC 1361  
  
Db 185 ttgcaggactgcaggttccacctcgtcggccgttcccttatttccagttgagctgtgca 244  
|||||  
Qy 1362 TTGCAGGAGCTGCAGGTTCCACCTCGCTCGCCCGCTCTTCATTTACGTTGAGCTGTGCA 1421  
  
Db 245 ggaagcgaagaacctctctgaagcattttcgtatccgtgacttgggaagagagctcacc 304  
|||||  
Qy 1422 GAGAAGCAGAAGACCTTCTGAGCATTTCGATTCGCTGACTTGGAGAGGATCCGTACC 1481  
  
Db 305 tcccttgtaatccccgtaagctgattgctatagccagatccccgcgcttcagacatcc 364  
|||||  
Qy 1482 TCCCTGTATATCCCGTGAATGCTGCTATAGCAGTATCCCGCGCTTCAGACATCC 1541

Db 365 ctcaagtgaattctgacagccgaggtttaaaagatataaagatcatgggtgattccatc 424  
|||||  
Qy 1542 CTCAGTGGATTTCTGACAGCCGAGTTTAAAGATATAAAGATCATGGGCTATTCCATAC 1601  
  
Db 425 gcaccatagactataggtataactgtgtgggttggtctcaatcctgatgaatttctagcta 484  
|||||  
Qy 1602 GCACCATAGACTATAGGTATAGTGTGTGGGTGGCTCAATTCCTGATGAATTTCTAGACTA 1661  
  
Db 485 acttttctgacatccatgcaggggaactgtatttctgtggattctgacccattgcaggtac 544  
|||||  
Qy 1662 ACTTTTCTGACATCCATGCAGGGGAATCTATTTTGTGGATTCTGACCATTCGAGGATC 1721  
  
Db 545 acaatgtataatgattcccaagtgagatcttttccagttgttgatgccttgagttt 604  
|||||  
Qy 1722 ACAATATGTAATATGATTTCCCAAGGTGGAGATCTTTCCAGCTTGTGATGCTTGAGTTT 1781  
  
Db 605 tgcacaacctgagtgagcaaatgtgtgtgtcctccctccagctggtgagaggaggttag 664  
|||||  
Qy 1782 TGCCAACCTGAGTGGCAATGTGATGTGCTCCCTCCAGCTGGTGAGAGGAGGATTAG 1841  
  
Db 665 agctgtgctgtttgtgattaccataatattggaagcagctgagggctagttaatccaa 724  
|||||  
Qy 1842 AGCTGTGCTGTTTGTGATTACCCATAATATTGGAGCAGCTGAGGCTAGTTAATCCAA 1901  
  
Db 725 acatgcatacaaatgtgctgagatatagtacagcccaaccttttctgttagcttt 784  
|||||  
Qy 1902 ACATGCATCAACAATTTGGCTCAGAAATGTAAGTATGTAAGCAACCTTTTCGTTAGTCTTT 1961  
  
Db 785 attaaaattataatgtgaattgacaggttttttttaatttccctctttttaaacc 844  
|||||  
Qy 1962 ATTAAATTAATATGTAATTTGGACAGCTTTTATTTTAAATTCCTCTTTTAAAC 2021  
  
Db 845 agttacgcttattactgaataataacaaagcaaaactcaagttatgcatcactt 904  
|||||  
Qy 2022 AGTTACGCTTATTACTGTAATAATACAAAGCAAACTCAAGTTATGTGATACCTT 2081  
  
Db 905 tggatacgaagaccataacataataacaaacataacattatacaaaagaatacttcat 964  
|||||  
Qy 2082 TGCATACGAAGACCATACATAATAACCAACATAACATATACACAAAGATACTTTTCA 2141  
  
Db 965 tatttgtgaattagtgcatcttcaaaagtaatacatatatacaaaactaggcaccacta 1024  
|||||  
Qy 2142 TATTGTGGAATTTAGTGCATTTCAAAAAGTAATCATATATCAAACTAGGACCACTA 2201  
  
Db 1025 agtctgtgattttgtttataatttaataatctcttatgagccctatatattcaaa 1084  
|||||  
Qy 2202 AGTCTGCTATTATTGTTATTAATTTAATTAATATATCTATGAGCCCTATATATTCAA 2261  
  
Db 1085 atattatgtaacatgaatccatgtttctttttcc 1120  
|||||  
Qy 2262 ATATTATGTTAAACATGTAATCCATGTGTTCTTTTTC 2297

RESULT 4  
LOCUS HUMIDS 36845 bp DNA PRI 15-AUG-1994  
DEFINITION Homo sapiens iduronate sulphate sulphatase (IDS) gene, complete cds.  
ACCESSION L35485  
KEYWORDS iduronate sulphate sulphatase.  
SOURCE Homo sapiens DNA.  
ORGANISM Homo sapiens  
Eukaryota; Animalia; Chordata; Vertebrata; Mammalia; Theria;  
Eutheria; Primates; Haplorhini; Catarrhini; Hominoidea.



REFERENCE 1 (bases 1 to 36845)  
AUTHORS Lu, F., Lu, J., Clingan, R.L., Wentland, M.A., Muzny, D.M., Gu, Y., Nelson, D.L. and Gibbs, R.A.  
TITLE Complete DNA sequence of the human iduronate sulphate sulphatase (ids) locus  
JOURNAL Unpublished (1994)  
COMMENT NCBI gi: 530140  
FEATURES  
source  
Location/Qualifiers  
1..36845  
/organism="Homo sapiens"  
/sequenced\_mol="DNA"  
<1057..1159  
/gene="IDS"  
/map="Xq27.3-q28"  
/number=1  
/codon\_start=1  
Join(1057..1159,1901..2037,2704..2881,5155..5243,7885..8085,9676..9846,15752..15878,19094..19267,22974..23446)  
/gene="IDS"  
/map="Xq27.3-q28"  
/notes="NCBI gi: 530141"  
/codon\_start=1  
/product="iduronate sulphate sulphatase"  
/translation="MPPRTGGLLGLVLSVCVALGSETQANSTTDALNVLIVDDLRLGCGKLUKLRSPNIDQLASHLLFQNAFAQAVCAF SRVSFLTCRRPDTTRL YDNSTYRHAGNESTIPQYFKNGYVTSVGKVFHPGICSSNHTDDSPYSWSPFPYH SSEKYENTKCRGDDGLHANLLCPVDVLDVPEGLDPKOSTEQAIOLLEKMTSASP FFLAVGYHKPHIPRYPKEFKQLYPLENITLAPDPEVDGLPPVAYNPWMDIRQEDV QALNIVPYGCPIDYFQRIKRGYSFASVYLDTVQGRLLSALDDIQLANSTIIAFTSD HGNALGEGENAKSNFNVATHVPLIFVYPGRTASLPEAGEKLFYLPDPEDSASQIME PGRQMDVLVSLFPIFLAGLAGLVPPPCVPVSHVELCREGNLLKHFRRDLEED PYPGPNPRELIAYSQYPRPSDIPQWNSDKP SILKIKIMGYSIRTIDYRYTVMVGPNPD EFLANFSDIHAGELFYVDSDP LQDRHNYNDSQGDLPQLIMP"  
1160..1900  
/gene="IDS"  
/map="Xq27.3-q28"  
/number=1  
1901..2037  
/gene="IDS"  
/map="Xq27.3-q28"  
/number=2  
/codon\_start=1  
2038..2703  
/gene="IDS"  
/map="Xq27.3-q28"  
/number=2  
2704..2881  
/gene="IDS"  
/map="Xq27.3-q28"  
/number=3  
/codon\_start=1  
2882..5154  
/gene="IDS"  
/map="Xq27.3-q28"  
/number=3  
5155..5243  
/gene="IDS"  
/map="Xq27.3-q28"  
/number=4  
/codon\_start=1  
5244..7884  
/gene="IDS"  
/map="Xq27.3-q28"

repeat\_region  
6827..7114  
/rpt\_family="Alu"  
/number=4  
repeat\_region  
7117..7647  
/rpt\_family="Line1"  
7885..8085  
/gene="IDS"  
/map="Xq27.3-q28"  
/number=5  
/codon\_start=1  
8086..9675  
/gene="IDS"  
/map="Xq27.3-q28"  
/number=5  
9676..9846  
/gene="IDS"  
/map="Xq27.3-q28"  
/number=6  
/codon\_start=1  
9847..15751  
/gene="IDS"  
/map="Xq27.3-q28"  
/number=6  
15752..15878  
/gene="IDS"  
/map="Xq27.3-q28"  
/number=7  
/codon\_start=1  
15879..19093  
/gene="IDS"  
/map="Xq27.3-q28"  
/number=7  
19094..19267  
/gene="IDS"  
/map="Xq27.3-q28"  
/number=8  
19268..22972  
/gene="IDS"  
/map="Xq27.3-q28"  
/number=8  
repeat\_region  
21573..21864  
/rpt\_family="Alu"  
repeat\_region  
21867..22051  
/rpt\_family="Line1"  
repeat\_region  
22098..22267  
/rpt\_family="Line1"  
22297..23446  
/gene="IDS"  
/map="Xq27.3-q28"  
/number=9  
25413..25686  
/rpt\_family="Alu"  
repeat\_region  
26878..27162  
/rpt\_family="Alu"  
repeat\_region  
29858..30098  
/rpt\_family="Alu"  
repeat\_region  
30767..31068  
/rpt\_family="Alu"  
repeat\_region  
33970..34090  
/rpt\_family="Alu"  
BASE COUNT 9117 a 8696 c 8303 g 10729 t  
ORIGIN

DB 107; Score 995; Match 100.0%; QryMatch 43.3%; Pred. No. 0.00e+00;  
Matches 995; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Db 22971 caggcaggaacatccatggaccccttggaacctgtgtctcttttccccacgctggctggac 23030  
|||||  
Qy 1302 CAGCGAGCAATCCATGGACCTTGTGGAATGCTGCTCTTTTCCAGCGCTGGCTGGAC 1361  
|||||  
Db 23031 ttgcaggaactgcaggttccacccctgcgccttccttcatttcacgttgagctgtgca 23090  
|||||  
Qy 1362 TTGCAGAGCTGCGAGGTTCACACTCGCTGCCCGCTTCCTTCATTTACGCTGACGTGCGCA 1421  
|||||  
Db 23091 gagaagcaagaacctctgaagcatttcgattccgtgacttgaagagatccgtacc 23150  
|||||  
Qy 1422 GAGAAGCAAGAACCTTCTCAAGCAATTTTCGATTCCTGCACTTGGAAAGGATCCGTACC 1481  
|||||  
Db 23151 tccctggtaatccccggtaactgattgcctatagccagatccccggccttcagacatcc 23210  
|||||  
Qy 1482 TCCCTGCTAATCCCGTCAACTGATTGCTATAGCCAGTATCCCGGCTTCAGACATCC 1541  
|||||  
Db 23211 etcagtggaattctgacaagccgagtttaaaagataaaagatcatgggtattccatc 23270  
|||||  
Qy 1542 CTCAGTGAATCTGCAACGCCAGCTTAAGATATTAAGATCATGGCTATTCCATAC 1601  
|||||  
Db 23271 gcaccatagaactataggtatctgtgtgggtggcttcaactcctgatgaattcttagcta 23330  
|||||  
Qy 1602 GCACCATAGACTATAGGTATGATCTGTGTGGGTGGCTTCAATCTCATGATTTCTAGCTA 1661  
|||||  
Db 23331 acttttctgaactccatgcagggaactgtatttgggattctgaaccttcagagatc 23390  
|||||  
Qy 1662 ACTTTTCTGACATCCATGCAGGGAACGTATTGTTGGAATTCGACCATTCGAGATC 1721  
|||||  
Db 23391 acaatatgtataatgattccccaggtggagatcttttccagttgttgatgccttagttt 23450  
|||||  
Qy 1722 ACAATATGTAATGATTCACCAAGCTGAGATCTTTTCCAGTTGTTGATGCTTGAGTTT 1781  
|||||  
Db 23451 tgcacaaccatggatggcaaatgtgtgtcccttccagctgggtgagagagaggttag 23510  
|||||  
Qy 1782 TCCCAACCATGATGGCAATGTGATGTGCTGCTCCCTCCAGCTGCTGAGAGGAGCTTAG 1841  
|||||  
Db 23511 agctggctgtttgtgattaccataatattggaagcagcctgaaggctagttaatccaa 23570  
|||||  
Qy 1842 AGCTGGTGTGTTTGTGATTACCCATAATATTGGAAGCAGCCTGAGGGCTAGTTAATCCAA 1901  
|||||  
Db 23571 acatgcatcaacaatttggcctgagaatattgtaacagccaaaccttttcgttagcttt 23630  
|||||  
Qy 1902 ACATGATCAACAATTTGGCTGAGATATGTAACAGCAAAACCTTTTCGTTAGTCTTT 1961  
|||||  
Db 23631 attaaattataatgttaattggaaccagtttttttaattccctctttttaaacc 23690  
|||||  
Qy 1962 ATTAATAATTTAATTTGGTAATTGGACAGCTTTTTTTTAAATTCCTCTCTTTTAAAC 2021  
|||||  
Db 23691 agttacggcttatttactgaataataacaagcaacaactcaagttagtcataacct 23750  
|||||  
Qy 2022 AGTTACGGCTTATTTACTGATTAATAACAAGCAACAACTCAAGTTATGTGTATACCTT 2081  
|||||  
Db 23751 tggatcgaagaccatacataataaccaacaataacattatcacaaagaataactttcat 23810  
|||||  
Qy 2082 TGGATACGAAGACCATACATAATAACAAACATAACATTATACAAAAGAAATACTTTTCAAT 2141  
|||||  
Db 23811 tatitgtgaatttagtgatttcaaaagttaattcatatatataaactaggcaccacata 23870  
|||||  
Qy 2142 TATTTTGGAAATTTAGTGCATTTTCAAAAAGTAATCATATATCAAACTAGGCACCACTA 2201  
|||||  
Db 23871 agttcctgattattttgttttaatttaataatatatcttatgagccctatatattcaaa 23930  
|||||

Qy 2202 ACTTCTCATTTTGTGTTTATAATTTAAATAATATCTTATGAGCCCTATATATTCAAA 2261  
|||||

Db 23931 attatattgtaaacatgaatccatgtttcttttc 23965  
|||||  
Qy 2262 ATATTATGTTAAACATGAATCCATGTTTCTTTTTC 2296  
|||||

RESULT 5  
LOCUS H06357 615 bp mRNA EST 21-JUN-1995  
DEFINITION Y179J02.r1 Homo sapiens cDNA clone 44414 5' similar to gb:M58342  
IDURONATE 2-SULFATASE PRECURSOR (HUMAN);.  
ACCESSION H06357  
KEYWORDS EST.  
SOURCE human clone=44414 library=Soares infant brain INIB vector=Lafmid BA

host=DH10B (ampicillin resistant) primer=M13RPI Rsite1=Not I  
Rsite2=Hind III Whole brain from a 73 days post natal female. 1st  
strand cDNA was primed with a Not I - oligo(dT) primer [5'  
AATCGAAGAAATTCGGCGCCCGCAGGAATTTTTTTTTTTT 3']; double-stranded  
cDNA was ligated to Hind III adaptors (Pharmacia), digested with  
Not I and directionally cloned into the Not I and Hind III sites of  
the Lafmid BA vector. Library went through one round of  
normalization. Library constructed by Bento Soares and M.Fatima  
Bonaldio.

ORGANISM Homo sapiens  
Eukaryota; Metazoa; Eumetazoa; Bilateria; Coelomata;  
Deuterostomia; Chordata; Vertebrata; Gnathostomata; Osteichthyes;  
Sarcopterygii; Choanata; Tetrapoda; Amniota; Mammalia; Theria;  
Eutheria; Archonta; Primates; Catarrhini; Hominoidea; Homo.  
REFERENCE 1 (bases 1 to 615)  
AUTHORS Hillier,L., Clark,N., Dubuque,T., Elliston,K., Hawkins,M.,  
Holman,M., Hultman,M., Kucaba,T., Le,M., Lemon,G., Marra,M.,  
Parsons,J., Rifkin,L., Rohlfing,T., Soares,M., Tan,F.,  
Trevisan,E., Waterston,R., Williamson,A., Wohlmann,P. and  
Wilson,R.  
TITLE The WashU-Merck EST Project  
JOURNAL Unpublished (1995)  
COMMENT

GDB: G00-416-955  
Contact: Wilson RK  
WashU-Merck EST Project  
Washington University School of Medicine  
4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108  
Tel: 314 286 1800  
Fax: 314 286 1810  
Email: est@watson.wustl.edu  
High quality sequence stops: 412  
Source: IMAGE Consortium, LLNL  
This clone is available royalty-free through LLNL ; contact the  
IMAGE Consortium (info@image.llnl.gov) for further information.

NCBI gi: 869909  
FEATURES  
source  
Location/Qualifiers  
1..615  
/organism="Homo sapiens"  
/clone="44414"  
/note="human"  
BASE COUNT 140 a 155 c 154 g 162 t 4 others  
ORIGIN

DB 142; Score 469; Match 93.5%; QryMatch 20.4%; Pred. No. 0.00e+00;  
Matches 575; Conservative 0; Mismatches 26; Indels 14; Gaps 14;  
Db 1 cttgcgtgnacaacagagcactgagcagccatcacagttgttggaagaagatgaaacgctc 60

[illegible]

RESULT 6  
ID H3357151 standard; RNA; EST; 615 BP.  
AC H06357;  
DT 23-JUN-1995 (Rel. 44, Created)  
DT 23-JUN-1995 (Rel. 44, Last updated, Version 1)  
DE U719H02.ri Homo sapiens cDNA clone 44414 5' similar to gb:M58342  
DE D109006 2-SULFATASE PRECURSOR (HUMAN).  
DE

OS *Homo sapiens* (human)  
OC Eukarya; Animalia; Metazoa; Chordata; Vertebrata; Mammalia;  
OC Theria; Eutheria; Primates; Haplorhini; Catarrhini; Hominoidea.  
RN [1]  
RP 1-615  
RA Hillier L., Clark N., Dubuque T., Elliston K., Hawkins M.,  
RA Holman M., Hultman M., Kucaba T., Le M., Lennon G., Marra M.,  
RA Parsons J., Rifkin L., Rohlfing T., Soares M., Tan F.,  
RA Trevasakis E., Waterston R., Williamson A., Wohlmann P., Wilson R.;  
RT "The WashU-Merck EST Project";  
RT Unpublished.  
CC GDB: G00-416-955 Contact: Wilson RK WashU-Merck EST Project

CC GDB: G00-416-955 Contact: Wilson RK WashU-Merck EST Protect

CC	Washington University School of Medicine 4444 Forest Park Parkway,
CC	Box 8501, St. Louis, MO 63108 Tel: 314 286 1800 Fax: 314 286 1810
CC	Email: est@watson.wustl.edu High quality sequence stops: 412
CC	Source: IMAGE Consortium, LML This clone is available royalty-free
CC	through LML ; contact the IMAGE Consortium (info@image.llnl.gov)
CC	for further information. NCBI gi: 869909
FH	Key                      Location/Qualifiers
FH	
FT	source                1..615
FT	/organism="Homo sapiens"
FT	/clone="44414"
FT	/note="human"
FT	
SQ	Sequence 615 BP; 140 A; 155 C; 154 G; 162 T; 4 other;

DB 3; Score	Matches 5/5; Conservative 0; Mismatches 26; Indels 14; Gaps 14;	DB 3; Score	Matches 5/5; Conservative 0; Mismatches 26; Indels 14; Gaps 14;
Db	1	cttgctgnacaaacagagcagcactgagcaagccatacagttgttgaaaagatgaaacagctc	60
Qy	709	CTTGCCCTGACAAACAGAGCACTGAGCAAGCATTACAGTTCTTTGGAAAGATGAAACGCTC	768
Db	61	agccagtcctttctctcctggccgttggtgatcataagccacacatcccccttcagataccc	120
Qy	769	AGCCAGTCCTTTCTCTCGCGGTTGGTGATCATTAAGCCACACATCCCTTCAGATACCC	828
Db	121	caaggaattcagaagtdgtatcccttgagacaatcacccctggccccgcagatcccgaggt	180
Qy	829	CNAGGAAITTCAGAGCTGTATCCCTTGGAGAACATCACCCCTGGCCCCCGCATCCCGAGGT	888
Db	181	ccctgatggcctaccocctgtggcctacaaacccctggatggacatcaggcaacgggaaga	240
Qy	889	CCCTGATGGCCTACCCCTCTGGGCTTACAACCCCTTGATGGACATCAGGCACACGGGAACA	948
Db	241	cgtccaaagccttaaacatcagtggtgcgtatggtccaattccctgtggactttcagcggaa	300
Qy	949	CGTCCAAAGCCTTAAACATCAGTGTGCGGTATGTCGCAATTCCTGTGACATTCCTCAGCGAA	1008
Db	301	aatccgcagagctactttgcctctgtgctatattggatatacacaggtcgccgcgcctctt	360
Qy	1009	AATCCGCAGAGCTACTTTGCGCTCTGTGTCAATTTGGATACACAGTGCGGCGGCTCTT	1068
Db	361	gagtgcttggacgatcttcagctgggccaaacagaccattcattgcaattacctcggt	420
Qy	1069	GAGTGTCTTGGACGATCTTCAGCTGG--CCAAACAGCACCAT-CATTGGATTTCCTCGGAT	1126
Db	421	tcatgggtggggttttagggtgaaacatgggaagatggcggaatggccaaatacacgaattttg	480
Qy	1127	-CATGGG-TGGGCTCT-AGG-TGAA-CATGG-AGANTGGG-CCAAATACAGCAATTTT-G	1178
Db	481	atgttggtttaccctggtttccctcgatatctatgctnctggagagcaggttcaatttcc	540
Qy	1179	ATGTCTTACCCATGTTT-CCCTCGATATTCTATGTCTCTGGAAAGCAGGCTTCACATT-CC	1236
Db	541	ggaggcaggcgagaagtttttttccctaacttgacctttttgattttccggttcacagttna	600
Qy	1237	GGAGGCGAGG-CGAGAAGCTTTTCCCTCTTACCTCGACCCCTTTTGATTTCGCGCTCACAGTTGA	1295
Db	601	ttggagcaggggaag	615
Qy	1296	T-GGAGCCAGGCAGG	1309

RESULT	7	S64545	1831 bp	mRNA	ROD	28-SEP-1993
LOCUS						

DEFINITION iduronate sulfatase [mice, thymus, mRNA Partial, 1831 nt].

ACCESSION S64545

KEYWORDS .

SOURCE mice thymus.

ORGANISM Mus sp.

Unclassified.

REFERENCE 1 (bases 1 to 1831)

AUTHORS Daniele,A., Russo,T., Ballabio,A. and Di Natale,P.

TITLE The mouse iduronate sulfatase gene: identification of a novel

transcript

JOURNAL Biochem. Biophys. Res. Commun. 194 (3), 1030-1037 (1993)

MEDLINE 93356764

REMARK GenBank staff at the National Library of Medicine created this

entry [NCBI gibseq 136230] from the original journal article.

This sequence comes from Fig. 1.

COMMENT NCBI gi: 404573

FEATURES Location/Qualifiers

source

1..1831

/organism="Mus sp."

/note="mice"

2..640

/partial

/note="bases 119-129 are left untranslated, three x'es are

filled in this region;

mismatches(23[Y->H],41[X->P],42[X->P],43[X->V]);

Description: iduronate sulfatase, IDS; Method: conceptual

translation supplied by author. This sequence comes from

Fig. 1. Author-given protein sequence differs from the

conceptual translation for reasons explained in citation.

IDS; NCBI gi: 404574"

/codon\_start=1

/product="iduronate sulfatase"

/translation="FLAVGYHRPHIPRYPKFQKLYPIENTILAPDPHPDLSLXXA

YNPMDIEREDVQALNISVPYGPIDPFQRIRQSFASVYLDYDVGHVLSALDDL

RLAHNTIIAFTSDHGALCEGEWAKYSNFDVATRPVLMLYPGRTAPLPAGQKLPF

YRPFDPASDMWDAGRHTEDELVELVSLFTLAGLGLPVLGAPSLIFMEFP"

BASE COUNT 494 a 375 c 444 g 518 t

ORIGIN

DB 117; Score 434; Match 85.0%; QueryMatch 18.9%; Pred. No. 0.00e+00;

Matches 533; Conservative 0; Mismatches 93; Indels 1; Gaps 1;

Db 1 ctctctggcagtggtaccacaagccgcgcatatcccttcagatcccccaaggaatttca 60

|||||

Qy 781 CTCTCTGGCGGTGGGTATCATACCCACACATCCCTTCAGATACCCAGGATTTCA 840

Db 61 gaagtgcacctgtgaaacataaacctggctcctgtatcccatgtctctatagctt 120

|||||

Qy 841 GAAGTGTATCCCTTGGAGACATCACCTGGCCCCCGATCCGAGGTCCTGATGGCCT 900

Db 121 accactgttagctacaacccctggatgatcatcaggagagagggaagtgtccaagcctt 180

|||||

Qy 901 ACCCCCTGTGGCTTACAAACCCCTGGATGATCATCAGGCAAGCAAGACGTCCAGCCTT 960

Db 181 aaacatcagtgctcctatggaccaattcctcgtgagatttcacgggaaatccgcgcagag 240

|||||

Qy 961 AAACATCAGTGTGGGTATGTGTCATTTCTGTGGACTTTTCAGCGAAATCCGCCAGAG 1020

Db 241 ctactttgtctgtcctattgtatctcaggtcgagacatgttttgatgctttgga 300

|||||

Qy 1021 CTACTTTGCTCTGTGTCATATTTTGATACACAGGTGGCGGCGCTCTTGAGTGTCTTGA 1080

Db 301 tgatcttcgtctagcccacacacattattgctttcacatctgatcacgagtgagcct 360

|||||

Qy 1081 CGATCTTCAGCTGGCCCAACAGACACCATCATTTGATTTACCTCGGATCATGGGTGGGCTCT 1140

Db 361 aggtgaacacggaatggccaagatcacagcaactttgtatgttgcacccgtgtgcctt 420

|||||

Qy 1141 AGGTGAACATGGCAATGGGCCAATACAGCAATTTTCATGTTGCTACCCATGTTCCCT 1200

Db 421 gatgcttatgtccctggaaggaagggcccaacttctcgggagggccagaagctctttcc 480

|||||

Qy 1201 GATATTCATGTTCTGTGAGGAGCGGCTTCACTTCGCGAGCGCAGGAGGCTTTTCCC 1260

Db 481 ataccgggacccttggatccggcctcggtggtgatggatgcaggcagccacacogagga 540

|||||

Qy 1261 TTACTCTGACCCCTTTTTCATTCGCGCTCACAGTTGATGAGCGCAGGCAATTCATGGA 1320

Db 541 cctcgtgaactgtatctctctccacacgctcgtcgtgacttgcaggactgccagt-cc 599

|||||

Qy 1321 CTTGTGGAACCTTGTCCTCTTTTCCAGCGTGGCTGCACTTGAGGACTGCAGGTTCC 1380

Db 600 tctcgtggtgcccccatccctcttttca 626

|||||

Qy 1381 ACCTCGCTGCCCCGCTTCCTTCATTCA 1407

RESULT 8

ID H5179158 standard; RNA; EST; 469 BP.

AC H05179;

DT 22-JUN-1995 (Rel. 44, Created)

DT 22-JUN-1995 (Rel. 44, Last updated, Version 1)

DE y185c04.r1 Homo sapiens cDNA clone 45095 5' similar to gb:M58342

DE IDURONATE 2-SULFATASE PRECURSOR (HUMAN);.

KW EST.

OS Homo sapiens (human)

OC Eukaryota; Animalia; Metazoa; Chordata; Vertebrata; Mammalia;

OC Theria; Eutheria; Primates; Haplorhini; Catarrhini; Homidae.

RN [1]

RP 1-469

RA Hillier L., Clark N., Dubuque T., Elliston K., Hawkins M.,

RA Holman M., Hultman M., Kucaba T., Le M., Lennon G., Marra M.,

RA Parsons J., Rifkin L., Rohlfing T., Soares M., Tan F.,

RA Trevasaki E., Waterston R., Williamson A., Wohlmann P., Wilson R.;

RT "The WashU-Merck EST Project";

RL Unpublished.

CC GDB: G00-417-636 Contact: Wilson RK WashU-Merck EST Project

CC Washington University School of Medicine 4444 Forest Park Parkway,

CC Box 8501, St. Louis, MO 63108 Tel: 314 286 1800 Fax: 314 286 1810

CC Email: est@wustl.edu High quality sequence stops: 103

CC Source: IMAGE Consortium, LNL This clone is available royalty-free

CC through LNL; contact the IMAGE Consortium (info@image.llnl.gov)

CC for further information. NCBI gi: 868731

FH Key Location/Qualifiers

FH source

FT source 1..469

/organism="Homo sapiens"

/clone="45095"

/note="human"

SQ Sequence 469 BP; 102 A; 128 C; 119 G; 116 T; 4 other;

DB 2; Score 412; Match 96.6%; QueryMatch 17.9%; Pred. No. 0.00e+00;

Matches 450; Conservative 0; Mismatches 11; Indels 5; Gaps 5;

Db 1 gaacataccctggccccgagctccctgatggctaccctctggtgacctaacaa 60

|||||

Qy 859 GAACATCACCTGGCCCCGATCCCGAGGTCCTGATGGCTTACCCCTGTGGCTTACA 918

Db 61 cccctggatgacatcaggcaacggaagacgtccaagccttaaacatcagtgccgta 120  
|||||  
Qy 919 CCCCTGATCGACATCAGGCAACGGGAAGACGTCCAAGCCTTAAACATCAGCTGCGCGTA 978  
  
Db 121 tggctcaattcctggaacttcagcgaaatccgacagactacttgctctgtgc 180  
|||||  
Qy 979 TGGTCCAAATTCCTGTGACATTCACGGGAAATCCGCCACAGACTATTGGCTCTGTCTG 1038  
  
Db 181 atatttgatcacacaggtcgccgcctcttgagtgcttggacgatcttcagctggccaa 240  
|||||  
Qy 1039 ATATTGATACACAGTCGCGCGCTCTTGAGTCTTTGGAGCATCTTCAGCTGGCCAA 1098  
  
Db 241 cagcaccatcattgatttacctcgatcatgggtgggtcctcagtggaacatggagaat 300  
|||||  
Qy 1099 CAGCACCATCATTTACCTCGGATCATGGGTGGCTCTAGGTGAACATGG-AGAA 1157  
  
Db 301 gggccaaatcacgaatttgatgttncatccatggtccctccatgattctatgttctcg 360  
|||||  
Qy 1158 GGGCAAATACAGCAATTTTGATGTGTGTACCCATGTTCCCTGTATTTCTATGTTCTGTG 1217  
  
Db 361 gaagcagcgtctcaactccgagcaggcgagaaggttttcccttacctngacctttt 420  
|||||  
Qy 1218 GAAGCAGCGTTCATCTCCGAGGACAGG-CGAGAGCTTTTCCCTTACTCGACCCCTTTT 1276  
  
Db 421 tgattcngcttaacagtgtgagcagcagggaaggaattccng 466  
|||||  
Qy 1277 -GATTCC-GCCTCACAGTTCATGG-AGCCAGGCGCAATCAATG 1319

RESULT 9  
LOCUS H05179 469 bp mRNA EST 21-JUN-1995  
DEFINITION y185c04.r1 Homo sapiens cDNA clone 45095 5' similar to gb:M58342  
IDRONATE 2-SULFATASE PRECURSOR (HUMAN) ;  
ACCESSION H05179  
KEYWORDS EST.  
SOURCE human clone=45095 library=Soares infant brain LNB1B vector=laflmid BA  
host=DH10B (ampicillin resistant) primer=M3R1 Rsite1=Not I  
Reite2=Hind III Whole brain from a 73 days post natal female. 1st  
strand cDNA was primed with a Not I - oligo (dT) primer [5'  
AACTCGAAGAAATTCGCGCCGACGAAATTTTTTTTTTTT 3']; double-stranded  
cDNA was ligated to Hind III adaptors (Pharmacia), digested with  
Not I and directionally cloned into the Not I and Hind III sites of  
the laflmid BA vector. Library went through one round of  
normalization. Library constructed by Bento Soares and M.Fatima  
Bonafide.

ORGANISM Homo sapiens  
Eukaryotae; Metazoa; Eumetazoa; Bilateria; Coelomata;  
Deuterostomia; Chordata; Vertebrata; Gnathostomata; Osteichthyes;  
Sarcopterygii; Chonata; Tetrapoda; Amniota; Mammalia; Theria;  
Eutheria; Archonta; Primates; Catarrhini; Homnidae; Homo.  
1 (bases 1 to 469)

REFERENCE  
AUTHORS Hillier,L., Clark,N., Dubuque,T., Elliston,K., Hawkins,M.,  
Holman,M., Hultman,M., Kucaba,T., Le,M., Lannon,G., Marra,M.,  
Parsons,J., Rifkin,L., Rohlfing,T., Soares,M., Tan,F.,  
Trevisakis,E., Waterston,R., Williamson,A., Wohlmann,P. and  
Willson,R.  
TITLE The WashU-Merck EST Project  
JOURNAL Unpublished (1995)  
COMMENT

GOB: G00-417-636  
Contact: Wilson RK  
WashU-Merck EST Project  
Washington University School of Medicine  
4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108

Tel: 314 286 1800  
Fax: 314 286 1810  
Email: est@watson.wustl.edu  
High quality sequence steps: 103  
Source: IMAGE Consortium, LINL  
This clone is available royalty-free through LINL; contact the  
IMAGE Consortium (info@image.lnl.gov) for further information.

NCBI gi: 868731  
FEATURES  
source  
Location/Qualifiers  
1..469  
/organism="Homo sapiens"  
/clone="45095"  
/note="human"  
BASE COUNT 102 a 128 c 119 g 116 t 4 others  
ORIGIN  
DB 142; Score 412; Match 96.6%; QryMatch 17.9%; Pred. No. 0.00e+00;  
Matches 450; Conservative 0; Mismatches 11; Indels 5; Gaps 5;

Db 1 gaacatcacctggccccgatcccgaggtccctgatggtccctaccctgtggcctacaa 60  
|||||  
Qy 859 GAACATCACCTGGCCCCGATCCCGAGTCCCTGATGGCTTACCCCTGTGGCTTACAA 918  
  
Db 61 cccctggatgacatcaggcaacggaagcgtccaagccttaaacatcagtgccgta 120  
|||||  
Qy 919 CCCCTGATCGACATCAGGCAACGGGAAGACGTCCAAGCCTTAAACATCAGCTGCGCGTA 978  
  
Db 121 tggctcaattcctgtggaatttcagcgaaatccgacagctacttgctctgtgc 180  
|||||  
Qy 979 TGGTCCAAATTCCTGTGACATTCACGGGAAATCCGCCACAGACTATTGGCTCTGTCTG 1038  
  
Db 181 atatttgatcacacaggtcgccgcctcttgagtgcttggacgatcttcagctggccaa 240  
|||||  
Qy 1039 ATATTGATACACAGGTTCGCGCGCTCTTGAGTCTTTGGACGATCTTCAGCTGGCCAA 1098  
  
Db 241 cagcaccatcattgatttacctcgatcatgggtgggtcctcagtggaacatggagaat 300  
|||||  
Qy 1099 CAGCACCATCATTTACCTCGGATCATGGGTGGCTCTAGGTGAACATGG-AGAA 1157  
  
Db 301 gggccaaatcacgaatttgatgttncatccatggtccctccatgattctatgttctcg 360  
|||||  
Qy 1158 GGGCAAATACAGCAATTTTGATGTGTGTACCCATGTTCCCTGTATTTCTATGTTCTGTG 1217  
  
Db 361 gaagcagcgtctcaactccgagcaggcgagaaggttttcccttacctngacctttt 420  
|||||  
Qy 1218 GAAGCAGCGTTCATCTCCGAGGACAGG-CGAGAGCTTTTCCCTTACTCGACCCCTTTT 1276  
  
Db 421 tgattcngcttaacagtgtgagcagcagggaaggaattccng 466  
|||||  
Qy 1277 -GATTCC-GCCTCACAGTTCATGG-AGCCAGGCGCAATCAATG 1319

RESULT 10  
LOCUS T08577 378 bp mRNA EST 03-AUG-1993  
DEFINITION EST06469 Homo sapiens cDNA clone H18650 5' end similar to  
Iduronate 2-sulfatase.  
ACCESSION T08577  
KEYWORDS EST.  
SOURCE Human clone=H18650 library=Infant brain, Bento Soares primer=M13  
Reverse.  
ORGANISM Homo sapiens  
Eucaryotae; Metazoa; Chordata; Vertebrata; Gnathostomata; Mammalia;  
Eutheria; Primates; Catarrhini; Hominidae; Homo.

REFERENCE 1 (bases 1 to 378)  
ADAMS, M.D., SOARES, M.B., KERLAVAGE, A.R., FIELDS, C. and VENTER, J.C.  
Rapid cDNA sequencing (expressed sequence tags) from a  
directionally cloned human infant brain cDNA library  
Nature Genet. 4, 373-380 (1993)

CONTACT: ADAMS, MD  
The Institute for Genomic Research  
932 Clopper Road, Gaithersburg, MD 20878  
Tel: 3018699056  
Fax: 3018699423  
Email: mdadams@tigr.org.

NCBI gi: 389605  
Location/Qualifiers  
source  
1..378  
/organism="Homo sapiens"  
/clone="HIBG50"  
/note="Human"

BASE COUNT 129 a 80 c 70 g 99 t  
ORIGIN

DB 56; Score 369; Match 99.5%; QryMatch 16.1%; Pred. No. 0.00e+00;  
Matches 376; Conservative 0; Mismatches 1; Indels 1; Gaps 1;

Db 2 gccgtactgttttaaaagagggaattaaaaaaa-cgtgtccattaccattata 60  
|||||  
|||||

Cp 2030 GCCGTAAGTGTTTAAAAAGAGGAATTAATAAACTGGTCCATTACCAATTATA 1971  
|||||

Db 61 aatttataaagactaaacgaagggttgctgttacatatctcaggccaaattgtt 120  
|||||

Cp 1970 AATTTTAATAGACTAAGCAAGGTTGGCTGTACATATTCAGGCCAAATTGTT 1911  
|||||

Db 121 gatcatgtttgattactagccctcaggctgttccaattattatggtaacacaaa 180  
|||||

Cp 1910 GATGCATGTTTGATTAACTAGCCCTCAGGCTGCTTCCAATATTATGGGTAAATCAAAA 1851  
|||||

Db 181 cagaccgtctaaactctctctcaccagctggaaggagacacatcacattgccatca 240  
|||||

Cp 1850 CGACCAGCTTAACCTCCTCTCACCAGCTGGAAGGAGCACATCACATTTGCCATCCA 1791  
|||||

Db 241 tggttgcaaaactcaaggcatcaacaactggaagatctccaccttgggaatcattat 300  
|||||

Cp 1790 TGGTTGGCAAACTCAGGCTCACAACTGGAAAGATCTCCACCCTTGGGAATCATAT 1731  
|||||

Db 301 acatattgtgatctgcaatgggtgcagaatccacaaaatacagttccctcgatgggtgt 360  
|||||

Cp 1730 ACATATTGTGATCTCGCAATGGGTGAGATCCACAAAATACAGTTCCTCGATGATGT 1671  
|||||

Db 361 cagaaaagttaactagaa 378  
|||||

Cp 1670 CAGAAAAGTTAGCTAGAA 1653  
|||||

RESULT 11  
LOCUS R15374 484 bp mRNA EST 13-APR-1995  
DEFINITION yf90e05.r1 Homo sapiens cDNA clone 29589 5' similar to gb:M58342  
IDUNONATE 2-SULFATASE PRECURSOR (HUMAN);  
ACCESSION R15374

KEYWORDS  
SOURCE human clone=29589 library=Soares infant brain 1NTB vector=Iafmid BA  
host=DH10B (ampicillin resistant) primer=M3RPI Reitel=Not I  
Reite2=Hind III Whole brain from a 73 days post natal female. 1st  
strand cDNA was primed with a Not I - oligo (dT) primer [5']

AACTGAAGAATTCGGCGCCGAGGAATTTTTTTTTTTT 3'; double-stranded  
cDNA was ligated to Hind III adaptors (Pharmacia), digested with  
Not I and directionally cloned into the Not I and Hind III sites of  
the Iafmid BA vector. Library went through one round of  
normalization. Library constructed by Bento Soares and M.Fatima  
Bonaldi.

ORGANISM Homo sapiens  
Eucaryotae; Metazoa; Chordata; Vertebrata; Gnathostomata; Mammalia;  
Eutheria; Primates; Catarrhini; Hominidae; Homo.

REFERENCE 1 (bases 1 to 484)

AUTHORS Hillier, L., Clark, N., Dubuque, T., Elliston, K., Hawkins, M.,  
Holman, M., Hultman, M., Kucaba, T., Le, M., Lennon, G., Marra, M.,  
Parsons, J., Rifkin, L., Rohlfing, T., Soares, M., Tan, F.,  
Trevasakis, E., Waterston, R., Williamson, A., Woldmann, P. and  
Wilson, R.

TITLE The WashU-Merck EST Project

JOURNAL Unpublished (1995)

COMMENT

GDB: G00-401-936

Contact: Wilson RK

WashU-Merck EST Project

Washington University School of Medicine

4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108

Tel: 314 286 1800

Fax: 314 286 1810

Email: est@wustl.wustl.edu

High quality sequence stops: 350

Source: IMAGE Consortium, LLNL

This clone is available royalty-free through LLNL; contact the  
IMAGE Consortium (info@image.llnl.gov) for further information.

NCBI gi: 769647

Location/Qualifiers

source

1..484

/organism="Homo sapiens"

/clone="29589"

/note="human"

BASE COUNT 112 a 128 c 121 g 114 t 9 others

ORIGIN

DB 38; Score 360; Match 97.8%; QryMatch 15.7%; Pred. No. 0.00e+00;  
Matches 402; Conservative 0; Mismatches 2; Indels 7; Gaps 7;

Db 1 ctgcnngagcaaacagagcactgagcaagccatacagttgttgaaaagatgaaacgt 60  
|||||

Qy 709 CTTGGCTCA-CAACAGAGCACTGAGCAAGCCATACAGTCTTGGAAAGATGAAACGT 767  
|||||

Db 61 cagccagtcctttcttctggcgttggtgtatcataagccacacatcccccttcagatacc 120  
|||||

Qy 768 CAGCCAGTCTCTTCTCTGGCGCTGGGTATCATAGCCACACATCCCTTCAGATACC 827  
|||||

Db 121 ccaaggaattcagaagttgtatcccttgagaaacatcacctggcccccgatcccgagg 180  
|||||

Qy 828 CAAGGAATTCAGAAATTGTATCCCTTGGAGAACATCACCTGGCCCCCGCATCCGAGG 887  
|||||

Db 181 tccctgatggcctaccctctggcctacaacccctggatggagcatcaggcaacgggaag 240  
|||||

Qy 888 TCCTCATGGCCCTTACCCCTGTGGCCTACAAACCCCTGGATGGACATCAGGCAACGGGAAG 947  
|||||

Db 241 aggtccaaagccttaaacatcagtggtgcgtatggtcccaattctctgtgactttcagcgga 300  
|||||

Qy 948 ACCTCAAGCTTTAAACATCAGTGTCCGTATGCTCAATTCTCTGTGGACTTTTCAGCGCA 1007  
|||||

Db 301 aaatccgcagagagttacttggctctgtgttctatttggatacacacaggtcgccgacct 360  
|||||

Qy 1008 AATCCGCCAGAGCT-ACITTCCTCTGTCT-CATATTTGGATACACAGCTGGCGCGCT 1065  
|||||  
Db 361 cttggagctcttgggacgatcttcagctgggccaacagcaccatcatg 411  
|||||  
Qy 1066 CTTC-AGTCCTTTGG-ACGATCTT-CAGCTGG-CCACAGCAGCATCATTC 1112  
|||||  
RESULT 12  
LOCUS H06300 439 bp mRNA EST 21-JUN-1995  
DEFINITION y179h02.s1 Homo sapiens cDNA clone 44414 3' similar to gb:M58342  
IDUONATE 2-SULFATASE PRECURSOR (HUMAN);  
ACCESSION H06300  
KEYWORDS EST.  
SOURCE human clone-44414 library=Soares infant brain LNB vector=Lafmid BA  
host=DH10B (ampicillin resistant) primer=Promega -2lml3 Reitel=Not  
I Rsite2=Hind III Whole brain from a 73 days post natal female. 1st  
strand cDNA was primed with a Not I - oligo(dT) primer [5'  
AACTGGAGATCCGCCCGCCGCGAGCAATTTTTTTTTTTT 3']; double-stranded  
cDNA was ligated to Hind III adaptors (Pharmacia), digested with  
Not I and directionally cloned into the Not I and Hind III sites of  
the Lafmid BA vector. Library went through one round of  
normalization. Library constructed by Bento Soares and M.Fatima  
Bonaldo.

ORGANISM Homo sapiens  
Eukaryota; Metazoa; Eumetazoa; Bilateria; Coelomata;  
Deuterostomia; Chordata; Vertebrata; Gnathostomata; Osteichthyes;  
Sarcopterygii; Chonata; Tetrapoda; Amniota; Mammalia; Theria;  
Eutheria; Archonta; Primates; Catarrhini; Hominoidea; Homo.  
REFERENCE 1 (bases 1 to 439)  
AUTHORS Hillier, L., Clark, N., Dubuque, T., Elliston, K., Hawkins, M.,  
Holman, M., Hultman, M., Kucaba, T., Le, M., Lennon, G., Marra, M.,  
Parsons, J., Rifkin, L., Rohlfing, T., Soares, M., Tan, F.,  
Trevaskis, E., Waterston, R., Williamson, A., Wohlmann, P. and  
Wilson, R.  
TITLE The WashU-Merck EST Project  
JOURNAL Unpublished (1995)  
COMMENT GDB: G00-416-955  
Contact: Wilson RK  
WashU-Merck EST Project  
Washington University School of Medicine  
4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108  
Tel: 314 286 1800  
Fax: 314 286 1810  
Email: est@watson.wustl.edu  
High quality sequence stops: 315  
Source: IMAGE Consortium, LNL  
This clone is available royalty-free through LNL; contact the  
IMAGE Consortium (info@image.llnl.gov) for further information.

NCBI gi: 869852  
Location/Qualifiers  
source 1..439  
/organism="Homo sapiens"  
/clone="44414"  
/note="human"  
BASE COUNT 135 a 83 c 83 g 133 t 5 others  
ORIGIN  
DB 142; Score 332; Match 96.1%; QryMatch 14.5%; Pred. No. 0.00e+00;  
Matches 398; Conservative 0; Mismatches 5; Indels 11; Gaps 11;  
Db 17 tttgttcttggattatttcagtaataagcgtaactnttttttaaaagagggaatt 76

Cp 2061 TTTGTTTGGCTTTGTTATTATTATTCAGTAAATAGCGTAACCTCTTTTAAAAACAGCGAAATT 2002  
|||||  
Db 77 aaaaaaaactggtccattaccattataaatttttaataaagactaaacgaaaggtt 136  
|||||  
Cp 2001 AAAAAAAGAACTGGTCCAAATACCAATATAAAATTTTAAATAAGACTAAACGAAAGGCTT 1942  
|||||  
Db 137 tggctgttacattctcagggccaaattnttgatgatctgtttgggattactagccctcag 196  
|||||  
Cp 1941 TGGCTGTGTACATATCTTCAGGCCAAATGTTGATGCATGTTTGGATTAACTAGCCCTCAG 1882  
|||||  
Db 197 gctgacttccaataattatgggttaatacacaacgaccagctcttaactctctctcaacca 256  
|||||  
Cp 1881 GCTC-CTTCCAAATATTATGGGTAATCAAAAACGACGCTCTAACTCTCTCTCAACCA 1823  
|||||  
Db 257 gctggaaggagacacatcacattgcccattccatgggttgggcaaaaactcaaggggcaatcaa 316  
|||||  
Cp 1822 GCTGGAAGGAGCAGCATCACATTTGCCATCCATGG-TTGG-CAAAAACCTCAAGG-CATCAA 1766  
|||||  
Db 317 caactgggaaaaggatctccacttgggggaatcattacatatgtgatcctggcaatg 376  
|||||  
Cp 1765 CAACTGG-AAAAG-ATCTCCACCTTGGG-AATCATATACATATTGTGATCCTG-CAATG 1710  
|||||  
Db 377 gggtcagantccacaanaatacagttccctggcagtggtgntgtcagaaaagtttag 430  
|||||  
Cp 1709 GG-TCAGATATCCACAAAATAACAGTTCCTCCTG-CATGG-ATGTCAGAAAAGTTAG 1659  
|||||

RESULT 13  
ID HS300154 standard; RNA; EST; 439 BP.  
AC H06300;  
DT 23-JUN-1995 (Rel. 44, Created)  
DT 23-JUN-1995 (Rel. 44, Last updated, Version 1)  
DE y179h02.s1 Homo sapiens cDNA clone 44414 3' similar to gb:M58342  
DE IDUONATE 2-SULFATASE PRECURSOR (HUMAN);  
KW EST.  
OS Homo sapiens (human)  
OC Eukaryota; Animalia; Metazoa; Chordata; Vertebrata; Mammalia;  
OC Theria; Eutheria; Primates; Haplorhini; Catarrhini; Hominoidea.  
RN [1]  
RP 1-439  
RA Hillier, L., Clark, N., Dubuque, T., Elliston, K., Hawkins, M.,  
RA Holman, M., Hultman, M., Kucaba, T., Le, M., Lennon, G., Marra, M.,  
RA Parsons, J., Rifkin, L., Rohlfing, T., Soares, M., Tan, F.,  
RA Trevaskis, E., Waterston, R., Williamson, A., Wohlmann, P., Wilson, R.;  
RT "The WashU-Merck EST Project";  
RL Unpublished.  
CC GDB: G00-416-955 Contact: Wilson RK WashU-Merck EST Project  
CC Washington University School of Medicine 4444 Forest Park Parkway,  
CC Box 8501, St. Louis, MO 63108 Tel: 314 286 1800 Fax: 314 286 1810  
CC Email: est@watson.wustl.edu High quality sequence stops: 315  
CC Source: IMAGE Consortium, LNL This clone is available royalty-free  
CC through LNL; contact the IMAGE Consortium (info@image.llnl.gov)  
CC for further information. NCBI gi: 869852  
FH Key Location/Qualifiers  
FT source 1..439  
/organism="Homo sapiens"  
/clone="44414"  
/note="human"  
FT  
SQ Sequence 439 BP; 135 A; 83 C; 83 G; 133 T; 5 other;  
Matches 398; Conservative 0; Mismatches 5; Indels 11; Gaps 11;  
DB 3; Score 332; Match 96.1%; QryMatch 14.5%; Pred. No. 0.00e+00;  
Matches 398; Conservative 0; Mismatches 5; Indels 11; Gaps 11;



Db 17 ttgtgttctgtatttattcagtaataagcgttaactntttttaaagagaggaaatt 76  
|||||  
Cp 2061 TTTGTTTCTTTGTTATTTATTCAGTAATAGCCGTAACTGTTTTTAAAGACGGGAATTT 2002  
|||||  
Db 77 aaaaaaaactggtccaattaccattataaatttttaataagagataaacgaaaaagtt 136  
|||||  
Cp 2001 AAAAAAAACTGGTCCAATTACCAATTTAAATTTTAAATAAGACTAAACCAAAAGTT 1942  
|||||  
Db 137 tggctgttaacattctcagggccaaattnttgatgcatgttttgattgaactagccctcag 196  
|||||  
Cp 1941 TGGCTGTTACATATTCCTCAGGCCAAATTTGTTGATGCATGTTTGATTAACCTAGCCCTCAG 1882  
|||||  
Db 197 gctgactccaattatggtgaatcacaaaacgaccagctctaaactcctctcaca 256  
|||||  
Cp 1881 GCTG-CTTCANATATATGGTAAATCAGAAACGACGAGCTTAACCTCCTCTCACA 1823  
|||||  
Db 257 gctgaaaggagcacatcacatttgcattccatgggttgggcgaactcaagggcacaa 316  
|||||  
Cp 1822 GCTGAAGGAGCACAATCAATTCGATCCATGG-TTGG-CAAAACTCAAGG-CATCAA 1766  
|||||  
Db 317 caactggaaaaggatctccacttgggggaatcattatcacatattgtatcctggcaatg 376  
|||||  
Cp 1765 CAACTGG-AAAAG-ATCTCACCTTGGG-AATCAITATACATATTGTGATCTG-CAAATG 1710  
|||||  
Db 377 gggtcagantccacaanaatcacagttccctcggcatggngtgcagaaaagttag 430  
|||||  
Cp 1709 GG-TCAGATCCACAAAATACAGTTCCCTCG-CATGG-ATGTCAGAAAAGTTAG 1659  
|||||

RESULT 14  
LOCUS HSC0SC011 327 bp RNA EST 14-FEB-1995  
DEFINITION H. sapiens partial cDNA sequence; clone c-0ac01.  
ACCESSION F05959  
KEYWORDS partial cDNA sequence; transcribed sequence fragment.  
SOURCE human.  
ORGANISM Homo sapiens  
Eukaryotes; mitochondrial eukaryotes; Metazoa/Eumycota group;  
Metazoa; Eumetazoa; Bilateria; Coelomata; Deuterostomia; Chordata;  
Vertebrata; Gnathostomata; Osteichthyes; Sarcopterygii; Chonata;  
Tetrapoda; Amniota; Mammalia; Theria; Eutheria; Archonta; Primates;  
Carnivora; Hominidae; Homo.  
REFERENCE 1 (bases 1 to 327)  
AUTHORS Genexpress.  
TITLE Direct Submission  
JOURNAL Submitted (19-JAN-1995) to the EMBL/GenBank/DBJ databases.  
Genethon, B.P. 60, 91002 Evry Cedex France and Genetique  
Moleculaire et Biologie du developpement, CNRS UPR420 B.P. 8, 94801  
Villejuif Cedex France. E-mail: genexpress@genethon.fr  
REFERENCE 2 (bases 1 to 327)  
AUTHORS Genexpress.  
TITLE The Genexpress cDNA program  
JOURNAL Unpublished  
REFERENCE 3 (bases 1 to 327)  
AUTHORS Auffray, C., Behar, G., Boie, F., Boucher, C., da Silva, C.,  
Devignes, M.D., Duprat, S., Houllgatte, R., Jumeau, M.N., Lamy, B.,  
Lorenzo, F., Mitchell, H., Mariage-Samson, R., Pietu, G., Pouliot, Y.,  
Sebastien-Kabakthchis, C. and Tessier, A.  
INWAG: Integrated molecular analysis of the human genome and its  
expression  
JOURNAL C.R. Acad. Sci., III, Sci. Vie 318, 263-272 (1995)  
COMMENT Cloning method: total mRNA was oligo-(dT) primed and directionally  
cloned 5' -> 3' into the HindIII -> NotI sites of the lafmid BA  
vector;

Sequencing method: single read, full automatic;  
Primer: M13 reverse  
cDNA sequence collinear to mRNA  
Stretch removed: nothing  
Normalization method: Bento Soares, P.N.A.S. 91:9228-9232 (1994);  
Genexpress library\_idt: C;  
Genexpress\_sequence\_idt: ylc-0sc01.  
NCBI gi: 669775  
FEATURES  
source location/Qualifiers  
1..327  
/organism="Homo sapiens"  
/clone\_lib="normalized infant brain cDNA from B. Soares,  
Psychiatry Dept. Columbia University USA"  
/sex="female"  
/tissue\_type="total brain"  
/dev\_stage="3 months old"  
/isolate="muscular atrophy patient"  
BASE COUNT 75 a 95 c 75 g 78 t 4 others  
ORIGIN  
DB 26; Score 321; Match 98.5%; QryMatch 14.0%; Pred. No. 0.00e+00;  
Matches 322; Conservative 0; Mismatches 5; Indels 0; Gaps 0;  
Db 1 aaaaatgaaaacgtcagccagctcttnttctgcccgttggtgatcataagccacaca 60  
|||||  
Qy 753 AAAGATCAAAACGCTCAGCCAGTCTCTTTCTTCCTGGCGTGGGTATCATAAAGCCACACA 812  
|||||  
Db 61 tcccttcagatcaccccaaggacttcagaagttgtatcccttgaggagaacatcacctgg 120  
|||||  
Qy 813 TCCCTTCAGATACCCCAAGAAATTCAGAAAGTTGTATCCCTTGGAGAACATCACCTGG 872  
|||||  
Db 121 ccccgatcccgaggttcctgacctaccctctggtcctacaaacctggatggaca 180  
|||||  
Qy 873 CCCCAGATCCCGAGGTCTCGTGGCTACCCCTGTGGCTTACACCCCTGGATGGACA 932  
|||||  
Db 181 tcaggcaacgggaagcagtcacagccttaaacatcagtgccgtatggtccaatctctg 240  
|||||  
Qy 933 TAGGCAACGGGAGACGCTCAAGGCTTAAACATCAGTGTGCCGTATGGTCCAAATTCCTG 992  
|||||  
Db 241 tggacttcacgggaaaaatccnccagagactacttgctctgtgtcatatttgatacac 300  
|||||  
Qy 993 TGGACTTTACGGGAAAATCCGACAGTACTTTGGCTGTGTATATTGGATACAC 1052  
|||||  
Db 301 aggtcggcgcgtctcttgatgctttgg 327  
|||||  
Qy 1053 AGGTGGCGCCCTCTTGAGTGCTTTGG 1079  
|||||

RESULT 15  
LOCUS T71926 423 bp mRNA EST 16-MAR-1995  
DEFINITION ye07a10.r1 Homo sapiens cDNA clone 117018 5' similar to gb:M58342  
IDURONATE 2-SULFATASE PRECURSOR (HUMAN);  
ACCESSION T71926  
KEYWORDS EST.  
SOURCE human clone=117018 library=Stratagene lung (#937210)  
vector=pBluescript SK- host=SOLR cells (kanamycin resistant)  
primer=M13RPI Reitel=EcoRI Raite=XhoI Normal lung tissue from a 72  
year old male. Cloned unidirectionally. Primer: Oligo dT. Average  
insert size: 1.0 kb; Uni-ZAP XR Vector; 5' adaptor sequence:  
5'-GAATTTCGGCAGCAG-3'; 3' adaptor sequence:  
5'-CTCGAGTTTTTTTTTTTTTTTTTTT-3'.  
ORGANISM Homo sapiens  
Eucaryotae; Metazoa; Chordata; Vertebrata; Gnathostomata; Mammalia;

## REFERENCE

## AUTHORS

Eutheria; Primates; Catarrhini; Homidae; Homo.  
1 (bases 1 to 423)  
Hillier, L., Clark, N., Dubuque, T., Elliston, K., Hawkins, M.,  
Holman, M., Hultman, M., Kucaba, T., Le, M., Lennon, G., Marra, M.,  
Parsons, J., Rifkin, L., Rohlfing, T., Soares, M., Tan, F.,  
Trevasakis, E., Waterston, R., Williamson, A., Wohlmann, P. and  
Wilson, R.

## TITLE

The WashU-Merck EST Project

## JOURNAL

Unpublished (1995)

## COMMENT

Contact: Wilson RK  
WashU-Merck EST Project  
Washington University School of Medicine  
4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108  
Tel: 314 286 1800  
Fax: 314 286 1810  
Email: est@watson.wustl.edu  
High quality sequence stops: 294  
Source: IMAGE Consortium, LBNL  
This clone is available royalty-free through LBNL; contact the  
IMAGE Consortium (info@image.llnl.gov) for further information.

NCBI gi: 686447

## FEATURES

Location/Qualifiers

1..423

source

/organism="Homo sapiens"

/clone="117018"

/notes="human"

BASE COUNT

88 a 114 c 101 g 119 t 1 others

ORIGIN

DB 71; Score 319; Match 95.8%; QryMatch 13.9%; Pred. No. 0.00e+00;

Matches 406; Conservative 0; Mismatches 4; Indels 14; Gaps 14;

Db 1 attttagtgcaccatggtccctgatattctatgtctggaagcgggttcac 60

Qy 1173 ATTTGATGTTCCATGCTCCCTGATATTCATGTTCTGGAAGCAGCTTCAC 1232

Db 61 ttccggag-caggcagaagctttcccttaoctgcaccccttttgattccgcctcacgt 119

Qy 1233 TTCCGGAGCAGCGCAGACCTTTTCCTTTACCTCGACCCCTTTTGATTCCGCCCTCACAGT 1292

Db 120 tgatggagccaggcaggcaatccatgacacctgtggaactgtgtctctttttcccaagc 179

Qy 1293 TGATGGAGCCAGCGCAGCAATCCATGACCTGTGTGAACTTGTGTCTCTTTTCCGACGC 1352

Db 180 tggctgactgcaggactgcaggtccacctgcctgcctgccttccttcatttcacgttg 239

Qy 1353 TGGCTGCACTTCAGGACTCCAGACTCCACCTGCTGCCCCCTTCCTTCATTTCAGTTG 1412

Db 240 agctgtgcaggaagcagaacaccttctggaagcatttccgattccgacttgggaag 299

Qy 1413 ACCTGTGCAG-AGAAGGCAAGAACCTCTC-AAGCATTTTCGATTCCGTCACCTGG-AAG 1469

Db 300 aggatccgtacctccctgggtaaatcccccggtgaacttgattgctatagccagatc 359

Qy 1470 AGGATCCGTACCTCCCTGG-TAA-TCCCC-GTG-AACT-GATT-GCCTATAGCCAGTATC 1523

Db 360 cccgggcttcagacatccctcagtggaatnttgacaagccgaggttttaaaagatata 419

Qy 1524 CCCCC-CTTTCAGATCCCTCAGTGG-AATTCT-GACAGCCGAGTTT-AAAAGATATA 1579

Db 420 aaga 423

||||

Qy 1580 AAGA 1583

Search completed: Wed Oct 4 14:43:28 1995  
Job time : 3284 secs.

\*\*\*\*\*

WAP5064 (TM)

\*\*\*\*\*

Release 2.1D John F. Collins, Biocomputing Research Unit.  
Copyright (c) 1993, 1994, 1995 University of Edinburgh, U.K.  
Distribution rights by IntelliGenetics, Inc.

MPsrch\_nn n.a. - n.a. database search, using Smith-Waterman algorithm

Run on: Wed Oct 4 13:42:53 1995; MasPar time 126.71 Seconds  
925.108 Million cell updates/sec

Tabular output not generated.

Title: >US-08-345-212-1  
Description: (1:2297) from US08345212.seq

Perfect Score: 2297  
N.A. Sequence: 1 CGGCTGTGTGGCAGCTTT.....TAATCCATGTTTCTTTTCC 2297  
Comp: GCCGACACAGCGCTCAGAA.....ATTAGTACAGAAAGG

Scoring table: TABLE default  
Gap 6

Nmatch STD : Dbase 0; Query 0

Searched: 61539 seqs, 25515148 bases x 2

Database: n-geneseq  
1 n-gen1  
2 n-gen2  
3 n-gen3  
4 n-gen4  
5 n-gen5  
6 n-gen6  
7 n-gen7  
8 n-gen8  
9 n-gen9  
10 n-gen10  
11 n-gen11

Statistics: Mean 10.033; Variance 5.812; scale 1.726

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	ID	Description	Pred. No.
c 1	231	10.1	233	10	Q77329	Human genome fragment 4.48e-140
c 2	99	4.3	1047	2	Q10572	Human Natriuretic Pep 4.33e-47
c 3	72	3.1	1047	2	Q10572	Human Natriuretic Pep 4.36e-29
c 4	48	2.1	204	1	N81164	Base substituted E.co 6.20e-14

5	45	2.1	204	1	N81164	Base substituted E.co	6.20e-14
6	48	2.0	91	9	Q51746	Oligonucleotide probe	3.90e-12
c 7	40	1.7	91	9	Q51746	Oligonucleotide probe	3.30e-09
8	37	1.6	501	3	N50033	Sequence encoding new	1.68e-07
9	35	1.5	501	3	N50023	Sequence encoding new	2.19e-06
10	35	1.5	501	3	N50031	Sequence encoding new	2.19e-06
11	35	1.5	501	3	N50025	Sequence encoding new	2.19e-06
12	35	1.5	501	3	N50027	Sequence encoding new	2.19e-06
13	35	1.5	501	3	N50028	Sequence encoding new	2.19e-06
14	35	1.5	501	3	N50032	Sequence encoding new	2.19e-06
15	35	1.5	498	3	N50034	Sequence encoding new	2.19e-06
16	33	1.4	501	3	N50029	Sequence encoding new	2.70e-05
17	33	1.4	501	3	N50030	Sequence encoding new	2.70e-05
18	31	1.3	501	3	N50024	Sequence encoding new	3.15e-04
19	30	1.3	565	6	Q35072	HCV envelope region n	1.05e-03
20	30	1.3	501	3	N50026	Sequence encoding new	1.05e-03
21	25	1.1	3871	2	N71302	HSV-1 gB and surround	3.29e-01
22	25	1.1	39	7	Q51787	Mixed oligonucleotide	3.29e-01
c 23	24	1.0	5393	5	Q34611	Plasmid pR3.	9.71e-01
c 24	23	1.0	501	3	N50031	Sequence encoding new	2.80e+00
c 25	23	1.0	501	3	N50023	Sequence encoding new	2.80e+00
c 26	23	1.0	501	3	N50032	Sequence encoding new	2.80e+00
c 27	23	1.0	501	3	N50029	Sequence encoding new	2.80e+00
c 28	23	1.0	498	3	N50034	Sequence encoding new	2.80e+00
c 29	23	1.0	1950	1	Q04472	Human papilloma virus	2.80e+00
c 30	23	1.0	501	3	N50028	Sequence encoding new	2.80e+00
c 31	22	1.0	3191	1	Q04092	3.2 kb KpnI-PvuII fra	2.80e+00
c 32	22	1.0	7241	3	Q15140	Genomic clone LE-ACC2	7.82e+00
c 33	22	1.0	36	2	Q11195	Ballast Constituent c	7.82e+00
c 34	22	1.0	3582	8	Q50879	HsmGluRI DNA.	7.82e+00
c 35	22	1.0	39	7	Q51787	Mixed oligonucleotide	7.82e+00
c 36	22	1.0	363	8	Q59724	Human brain Expressed	7.82e+00
37	22	1.0	8920	11	Q62924	Carbamoyl-phosphate-s	7.82e+00
38	22	1.0	2504	6	Q36467	M. scrofulaceum shutt	7.82e+00
c 39	22	1.0	565	6	Q35072	HCV envelope region n	7.82e+00
c 40	21	0.9	4221	11	Q64046	A253 derived bcl-1 cD	2.12e+01
c 41	21	0.9	2463	7	Q43711	Sequence of the human	2.12e+01
c 42	21	0.9	5852	2	Q11710	Dictyostelium plasmid	2.12e+01
c 43	21	0.9	48	9	Q53202	CYCD1-H034 first poly	2.12e+01
c 44	21	0.9	48	9	Q53204	CYCD1-G068 first poly	2.12e+01
c 45	21	0.9	501	3	N50024	Sequence encoding new	2.12e+01

## ALIGNMENTS

RESULT 1  
ID Q77329 standard; DNA; 233 BP.  
AC Q77329;  
DT 15-NOV-1994 (first entry)  
DE Human genome fragment (Preferred).  
KW Brain; placenta; bone marrow; genetic analysis; gene mapping;  
KW detection; homology; human; adrenal tissue; ds.  
OS Homo sapiens.  
PN M09401548-A.  
PD 20-JAN-1994.  
PR 13-JUL-1993; G01467.  
PR 13-JUL-1992; GB-014857.  
PA (MEDI-) MEDICAL RES COUNCIL.  
PI Gross J, Hadfield KM, Howells D, Kelly M, Shaw D;  
PI Sibson DR, Starkey M;  
DR WFI; 94-035056/04.  
PT New nucleic acid fragment encoding gene products - can be used  
PT for genetic analysis and mapping  
PS Claim 1; Page 470; 616pp; English.



Result No.	Score	Query		DB	ID	Description	Pred. No.
		Match	Length				
1	172	4.2	556	9	R51355	Sequence of murine bo	2.75e+05
2	110	2.7	1671	1	R06341	Signal peptide and fi	1.74e+00
3	110	2.7	2721	2	P70647	Sequence of N-termina	1.74e+00
4	107	2.6	351	8	R44211	Caffeine demethylase.	2.85e+00
5	97	2.4	1301	5	R28304	Acid alpha-amylase en	1.42e+01
6	96	2.4	1074	4	R24102	Marek's disease virus	1.66e+01
7	92	2.3	817	9	R47201	DP42 mannosyl transfe	3.09e+01
8	90	2.2	455	7	R39344	Epic protein.	4.20e+01

## ALIGNMENTS

RESULT	1	
ID	R51355 standard; Protein; 556 AA.	
AC	R51355;	
DT	12-OCT-1994 (first entry)	
DE	Sequence of murine bone-related sulphatase-like precursor	
DE	protein OSF-8.	
KW	OSF-8; bone-related sulphatase-like protein; osteoporosis;	
KW	Paget's disease; osteomalacia; therapy; diagnosis.	
OS	Mus musculus.	
PN	AU9344921-A.	
PD	03-MAR-1994.	
PF	27-AUG-1993; 044921.	
PR	28-AUG-1992; JP-230030.	
PR	03-DEC-1992; JP-324034.	
PA	(FARH ) HOECHST JAPAN LTD.	
PA	(FARH ) HOECHST JAPAN KK.	
PI	Amann E, Ito T, Otawara-hamamoto Y, Takeshita S;	
DR	WPI; 94-109856/14.	
DR	N-PSDB; Q51355.	
PT	Bone-related sulphatase-like protein, OSF-8 - degrades sulphate	
PT	groups of proteoglycan sugar chains, useful for treating bone	
PT	metabolic diseases	
PT	Claim 2; Page 17-22; 29pp; English.	

CC cDNA of mouse OSF-8 (mOSF-8) was isolated from mouse osteoblastic  
 CC cell line MC3T3-E1 cDNA library by a combination of PCR and the  
 CC subtraction method, and by the differential screening method. OSF-8  
 CC has homology with known representative sulphatase molecules but  
 CC belongs to a new subclass different from those so far reported.  
 CC The inventors claim the precursor protein (R51355 AAs 1-556) which  
 CC includes the signal peptide (AAs 1-18), and they also claim a protein  
 CC comprised of residues 19-556. They also claim DNA or RNA encoding  
 CC the claimed proteins.  
 SQ Sequence 556 AA;

DB 9; Score 172; Match 37.2%; QryMatch 4.2%; Pred. No. 2.75e+05;  
 Matches 29; Conservative 19; Mismatches 28; Indels 2; Gaps 2;

Db 274 knirafyyamcaedamleallalhlkldllqtktiyvtedhg-emamehrqfkymsye 332

Qy 294 RKIRQSFASVSYLDTVGRLSALDDQLANSTIIAFTSDHGWLALG-ENGEMAKYSNFD 352

Db 333 asvhvpllmqgikanal 350

Qy 353 VATHVPLIFVPGRTASL 370

## RESULT 2

ID R06341 standard; protein; 1671 AA.

AC R06341;

DE 14-DEC-1990 (first entry)

DT Signal peptide and first 1643 AAs of mature apoB protein.

KW Atherosclerosis; lipid binding peptide; LBP; phospholipid;

KW Lecithin cholesterol acyl transferase; L-CAT; apo A1;

KW Chinese hamster ovaries;

OS Homo sapiens.

FH Key Location/Qualifiers

FT Peptide 1..27

FT /label=Signal peptide.

FT /label=Mature peptide.

PN US943527-A.

PD 24-JUL-1990.

PF 27-FEB-1986; 834300.

PR 04-OCT-1985; US-784418.

PR 04-DEC-1985; US-804692.

PR 27-FEB-1986; US-834300.

PA (CALB-) CALIFORNIA BIOTECHN.

PI Protter AA, Vigne JL, Mallory JB, Talmadge KD, Kane JP;

DR MPI; 90-246622/32.

DR N-PSDB; 005590.

PT Prodn. of purified lipid-binding peptide - capable of binding to

PT phospholipid(s), useful in e.g. atherosclerosis.

PS Disclosure; pp; English.

CC Lipid-binding peptides in conjunction with phospholipids are

CC effective in reversing cholesterol transport, and may enhance

CC lecithin cholesterol acetyl transferase, lowering the rate of

CC plaque formation and atherosclerosis.

SQ Sequence 1671 AA;

DB 1; Score 110; Match 27.3%; QryMatch 2.7%; Pred. No. 1.74e+00;  
 Matches 24; Conservative 22; Mismatches 37; Indels 5; Gaps 5;

Db 1292 nsaki-eipilpggksrdlkmletvrtpalhf-kvqfhlperfqvptftipklyqlq 1349

Qy 187 DVLDPVEGTLPDKQSTEQAIQLLEKMKTSASPFFLAVGYHKPHIPFRYPK-EFQKLYPLE 245

Db 1350 -vpilgvlstnvnynln-wsgllqw 1375

Qy 246 NITLAPDEVPDGLPPVAYNPWMDIRQR 273

## RESULT 3

ID P70647 standard; protein; 2721 AA.

AC P70647;

DT 17-APR-1991 (first entry)

DE Sequence of N-terminal apolipoprotein B (apoB).

KW Lipid-binding peptide; apolipoprotein purification.

OS Homo sapiens.

FH Key Location/Qualifiers

FT Peptide 87..167

FT /label=Signal

FT Protein 168..8247

PN M08702062-A.

PD 09-APR-1987.

PF 02-OCT-1986; 002075.

PR 04-OCT-1985; US-784418.

PR 04-DEC-1985; US-804692.

PR 27-FEB-1986; US-834300.

PA (BIOT-) BIOTECHN RES PARTN.

PI Protter AA, Vigne JL, Mallory JB, Talmadge KD;

DR MPI; 87-108703/15.

DR N-PSDB; N70996.

PT Purified lipid-binding peptide prodn. - by expression of

PT recombinant system and adding lipid to obtd. peptide mixt. to

PT form lipo-peptide complex

PS Claim 50; pp73-85; 123pp; English.

CC ApoB is a lipid-binding peptide adapted to bind to phospholipids at

CC one or more amphipathic alpha-helical peptide regions. ApoB gene is

CC pref. expressed in Chinese hamster ovary (CHO) cells and the gene is

CC introduced in an expression vector having a regulatable promoter

CC derived from the human metallothionein II gene.

SQ Sequence 2721 AA;

DB 2; Score 110; Match 27.3%; QryMatch 2.7%; Pred. No. 1.74e+00;  
 Matches 24; Conservative 22; Mismatches 37; Indels 5; Gaps 5;

Db 1292 nsaki-eipilpggksrdlkmletvrtpalhf-kvqfhlperfqvptftipklyqlq 1349

Qy 187 DVLDPVEGTLPDKQSTEQAIQLLEKMKTSASPFFLAVGYHKPHIPFRYPK-EFQKLYPLE 245

Db 1350 -vpilgvlstnvnynln-wsgllqw 1375

Qy 246 NITLAPDEVPDGLPPVAYNPWMDIRQR 273

## RESULT 4

ID R44211 standard; Protein; 351 AA.

AC R44211;

DT 26-MAY-1994 (first entry)

DE Caffeine demethylase.

KW Caffeine demethylase; CDM; Pseudomonas; 3-methyl-7-alkyl-xanthine;

KW 1,3-dimethyl-7-alkyl-xanthine; demethylation; ds.

OS Pseudomonas sp.

PN DE4316882-A.

PD 25-NOV-1993.

PF 19-MAY-1993; 316882.

PR 20-MAY-1992; JP-154380.

PR 27-OCT-1992; JP-312954.

PA (AMAN ) AMANO PHARM KK.

PI Imai Y, Koide Y, Nakane S;

DR MPI; 93-378610/48.

Result	Query			DB	ID	Description	Pred. No.
	No.	Score	Match Length				
1	4069	100.0	550	9	A36493	iduronate-2-sulfatas	0.00e+00
2	3129	76.9	563	10	A47153	iduronate-2-sulfatas	0.00e+00
3	1300	31.9	212	10	PN0565	iduronate-2-sulfatas	2.47e-196
4	213	5.2	583	9	A32641	steryl-sulfatase (EC	4.55e-14
5	211	5.2	567	9	A37362	arylsulfatase (EC 3.	8.76e-14
6	211	5.2	522	9	J01299	N-acetylglactosamin	8.76e-14

**KEYWORDS** sulfuric ester hydrolase  
**SUMMARY** #length 550 #molecular-weight 61872 #checksum 5247



DB 9; Score Matches 550; Conservative 0; Mismatches 0; Indels 0; Caps 0;	
Db	1 mppprtgrglwlgivlasvcvlgsetqanstdtdalnvlvliiddlrslgcygdklvr 60 
Qy	1 MPPPRTRGLLWLGIVLSSVCVALGSETQANSTTDALNVLIIIVDDLRLPSTLCYGDKLVR 60 
Db	61 spnidqlashellfqnaqaqvcaprsvfltr:pdtr:rlfdnysyrvhagnfatip 120 
Qy	61 SPNIDQLASHSLFQNAQAQVACAPRSVFLTRPDTRTRLFDNYSYRVHAGNFSTIP 120 
Db	121 qyfkengvytmsvgkvfhpgissnhtddpswafpyhpssekyentkcrpgdgelha 180 
Qy	121 QYFKENGVTYMSVGKVFHPGISSNHTDPSYMSFFPYHPSEKYENTKCRGPDGELHA 180 
Db	181 nllcpdvdlvdpogt:lpdk:qeteqaiqllekmttsaspfllavgyhkphiprpykfeqk 240 
Qy	181 NLLCPVDVLDVPEGTLPDKQTEQAQLLEKMTTSASPFLLAVGYHKPHIPRPYKFEQK 240 
Db	241 lyplenitlapdpevpdglppvaynpmdir:qredvqalnisvpygpipdfq:kirqsy 300 
Qy	241 LYPLENIITLAPDEVPDGLPPVAYNPMDIRQREDVQALNISVPGIPDFQKIRQSY 300 
Db	301 faevsyldtqvr:Llaalddqlan:tiiafstdhgwalghegwakynfdvathvpli 360 
Qy	301 FASVSTLDTVQGRALSALDDQLANSTIIATSDHGWALGEGHWAKYSNFDVATHVPLI 360 
Db	361 fyvprtasi:peageklfpyldpf:daeq:lmepqrqsm:dlvelvs:lfpt:laqlaglvpp 420 
Qy	361 FYVPRTASLPEAGEKLFYLDPFDSASQLEMPGRQSMDLVELVSLFPTLAGLAGLVPP 420 
Db	421 rcvpvsfhvelcregknlkhnfr:frdeedpy:lpgnpreliayeqyprpsdi:lpqwnsdxp 480 
Qy	421 RCPVPSFHVELCREGKNLKHFRFRDEEDPYLPGNPRELIAYSQYPRPSDIPQWNSDXP 480 
Db	481 slkdikimgysirtidyrtvwwgfnpdeflanf:sdhagelyfvdsdplqdhmnyndsq 540 
Qy	481 SLKDIKIMGYSIRTDYRTVWWGFNPDEFLANFSDIHAGELYFVDSDDLQDHMNYNDSQ 540 
Db	541 ggdifqgllmp 550 
Qy	541 GGDIFQLLMP 550 

```

2
RESULT
ENTRY
TITLE
ORGANISM
DATE
03-May-1994
A47153
A47153
Daniele, A.; Faust, C.J.; Herman, G.E.; Di Natale, P.;
Ballabio, A.
Genomics (1993) 16:755-757
#journal
#title
Cloning and characterization of the cDNA for the murine
iduronate sulfatase gene.
#cross-references M01D:93315172
#accession A47153
#status preliminary
#molecule_type mRNA
#residues 1-563
#label DAN
#cross-references GB:L07921

```

KEYWORDS	sulfuric ester hydrolase
SUMMARY	#length 563 #molecular-weight 63437 #checksum 8505
DB 10; Score	3129; Match 79.0%; QryHatch 76.9%; Pred. No. 0.00e+00;
Matches 433; Conservative 58; Mismatches 53; Indels 4; Gaps 4;	
Db	19 araiwrqlsfallgglcfaiatesaagqsatdaillliivddllrplsgcygdklvrsnp 78
Qy	4 PRTGRGLLWGLVSSVCVALGSETQANSTTDALNVLILLIIVDDLRLPSLGCYGDKLVRSNP 63
Db	79 idqlaehsvlfqmafagqacapsrsvflgrtpdtrtlydinsyrvvhgfnstipgyf 138
Qy	64 IDQLASHSLIQNAFAQAQACAPSRVSFLTGRDPTTRLTYDINSYRVVHAGFNSTIPQYF 123
Db	139 kengyytmsvgvkfhpqissnshsddypyswfspphyhssekyentktckgqdklhanll 198
Qy	124 KENGYYTMSVGKVFHPGFISSNHTDDSPYSWSEPPHYPSSEKYENTKTCRCPDGEJHANLL 183
Db	199 cpvadvapreptldkqsteeairllektatgspfflawgyhkhphipryktefcklyp 258
Qy	184 CPVDVLDVPECTLPDKQSTEQAIQLLEKWKTSAPFFLAVGYHKPHIPRYKPEFKLYP 243
Db	259 lenitlapdphvdsldpvyaynpwmdiredvqalnisyvgpipefgrkirsyfas 318
Qy	244 LENITLAPDPEVDGLPPVAYNPWMDIROEDVQALNISVPGPIPVDFQBKIRSYFAS 303
Db	319 vsyldtqgvhvisalddrlahntiaftsdhgwalghekwakysnfvdvtrvplmlyv 378
Qy	304 VSYLDTVQGRULSALDDIQLANSTTIAFTSDHGWALGHEKWAKYSNFDVATHVPLIFYV 363
Db	379 pgrtapipeaqckliffyrdpdpasdwmdaghtedvelvelftlaglaglpyllgap 438
Qy	364 PGRTAPIPEAGEKLIFFYLDPDFDSASQLMEPRGRQMSDVLVELSIFPTLAGLAGLQVPPRC 423
Db	439 sllfmlsfaeakar-ifrsic-smtwkretclvpreliayeqyprpadfpqwnsdktpl 496
Qy	424 VPSFHVLCREGKNLLKHFRRFDLE-EDFYLPGNPRELIAYSQYPRPSDIPQWNSDKPSL 482
Db	497 ndirswdiyth-vdyrytvwvfgdpeslanfsdihagelyfvdsdplqdnhyvndshg 555
Qy	483 KDIKIKWYSITIDYRYTVWVGFPNDFELANFSDIHAGELYFVDSPLQDNHYVNDSDGG 542
Db	556 glhsrlrp 563
Qy	543 DLQQLMPP 550

```

RESULT      3
ENTRY       iduronate-2-sulfatase (EC 3.1.6.13) (clone MTA13) - mouse
TITLE       (fragment)
ALTERNATE_NAMES
ORGANISM    iduronate sulfatase
             #formal name Mus musculus #common name house mouse
DATE        05-Aug-1994 #sequence_revision 05-Aug-1994 #text_change
             05-Aug-1994
ACCESSIONS  PN0565
REFERENCE   Daniele, A.; Russo, T.; Ballabio, A.; Di Natale, P.
             Biochem. Biophys. Res. Commun. (1993) 194:1030-1037.
             The mouse iduronate sulfatase gene: Identification of a novel
             transcript.
             #accession PN0565
             #molecule_type mRNA
             #residues 1-212 ##label DAN
             ##titles

```

##note the authors translated the codon CAT for residue 23 as Tyr

COMMENT This enzyme is a lysosomal hydrolase that splits the sulfate group from iduronic acid residues present in the mucopolysaccharides dermatan and heparan sulfate.

KEYWORDS sulfuric ester hydrolase

SUMMARY #length 212 #checksum 4241

DB 10; Score 1300; Match 86.9%; QryMatch 31.9%; Pred. No. 2,47e-196; Matches 173; Conservative 18; Mismatches 8; Indels 0; Gaps 0;

Db 1 flavghkphipfrykpfqklhplnitlapdhvpslppvaynpwmdireredval 60

Qy 220 FLAVGYKHPIFRYKPFQKLYPLENITLAPDPEVDGLPPVAYNPWMDIREREDVAL 279

Db 61 nlsvpygpipefqrkirsyfasvyltqvghvlsalddrlahntiiaftsdhgal 120

Qy 280 NISVPYGPVDFQKIRQSYFASVYLDTVQGRLLSALDDQLANSTIIAFTSDHGMAL 339

Db 121 gehewakysnfdatrvplmlyvggtpaplaaqqklfpyrdpfdpaswdmagrhted 180

Qy 340 GEHEWAKYSNFDATRVPLMLYVGGTPAPLAQAQKLFPHYRDPFDPASWDMAGRHTED 399

Db 181 lvelslftplaglaglpv 199

Qy 400 LVELSLFTPLAGLAGLQV 418

RESULT 4

ENTRY A32641 #type complete

TITLE steryl-sulfatase (EC 3.1.6.2) - human

ALTERNATE\_NAMES arylsulfatase C

ORGANISM #formal name Homo sapiens #common name man

DATE 21-May-1990 #sequence\_revision 21-May-1990 #text\_change 28-Apr-1993

ACCESSIONS A32641

REFERENCE A32641

authors Stein, C.; Hille, A.; Seidel, J.; Rijnbout, S.; Waheed, A.;

Schmidt, B.; Geuze, H.; von Figura, K.

#journal J. Biol. Chem. (1989) 264:13865-13872

#title Cloning and expression of human steroid-sulfatase. Membrane topology, glycosylation, and subcellular distribution in BHK-21 cells.

#cross-references MUID:89340479

#accession A32641

##molecule type mRNA

##residues 1-583 ##label STE

KEYWORDS glycoprotein; sulfuric ester hydrolase

FEATURE

1-21 #domain (or 1-23) signal sequence #status predicted

22-583 #product (or 24-583) steryl-sulfatase #status predicted

47,259 #label MAT

#binding\_site carbohydrate (Asn) (covalent) #status experimental

SUMMARY #length 583 #molecular-weight 65492 #checksum 7062

DB 9; Score 213; Match 27.6%; QryMatch 5.2%; Pred. No. 4.55e-14; Matches 92; Conservative 85; Mismatches 127; Indels 29; Gaps 24;

Db 18 eaeshaarp-nlilmaddlgiqpgcgnktrtpnidrlasgkvklthlaaeplct 76

Qy 27 ETQANSTTDALNVLIIIVDDL-RPSLCYCGKLVKSPNIDQLASHSLIFQNAFAQVCA 85

Db 77 paraafmtryvrgmaswstvgvflftassggpltdetfakllkdqgstaligk-w 135

Qy 86 PSRVSFILTCRRDPDTRLTYDFN-S--Y-WRVHAGNEST--I--PQYFKENGVTMTSGVKVF 137

Db 136 hlgmschsktdcfchphhghfn-yf-ygielnlndrckpgegevtfgkrlyrlplqliv 193

Qy 138 HPGTSSN-HTD--DSPYSWSFPPYHPSSKEYENTKTCRGPDDELHAN-LLCPVDV-LDVP 192

Db 194 gvtll-claalncilqlhvpqlgvffsallflaaliltflglfhyfrplnc-fmmrnyeli 251

Qy 193 EGTLPDQKSTEQATQILLE-KMKTSASFFFLAVGYHKPHIF-RYPKEFKLYPLENITIA 250

Db 252 qpmasydnltqcltveaaqfignrtetpflilvlylhvhtalfeskdfaqsgqvygda 311

Qy 251 PDPEVPDGLPP-VAYNPWMDI-RQRED--VQALN-ISVPYGP IPV-DFQKIRQOS-YFAS 303

Db 312 veemdwavvgqilnlldeirlandtliiyftsdqg 344

Qy 304 VSYLDTVQGRLLSALDDQLANSTIIAFTSDHG 336

RESULT 5

ENTRY A37362 #type complete

TITLE arylsulfatase (EC 3.1.6.1) precursor - sea urchin

ALTERNATE\_NAMES (Strongylocentrotus purpuratus)

ORGANISM #formal name Strongylocentrotus purpuratus #common\_name purple urchin

DATE 17-Apr-1993 #sequence\_revision 17-Apr-1993 #text\_change 17-Feb-1994

ACCESSIONS A37362

REFERENCE A37362

authors Yang, Q.; Angerer, L.M.; Angerer, R.C.

#journal Dev. Biol. (1989) 135:53-65

#title Structure and tissue-specific developmental expression of a sea urchin arylsulfatase gene.

#accession A37362

##status preliminary

##molecule\_type mRNA

##residues 1-567 ##label YAN

##cross-references GB:M25815

KEYWORDS sulfuric ester hydrolase

SUMMARY #length 567 #molecular-weight 62477 #checksum 214

DB 9; Score 211; Match 32.8%; QryMatch 5.2%; Pred. No. 8.76e-14; Matches 44; Conservative 33; Mismatches 48; Indels 9; Gaps 8;

Db 54 lqgtqghrtamtknvillladdmgvdlsvyghp-tqepqfidgmanglrftqgsgd 112

Qy 24 LGSETQANSTTDALNVLIIIVDDL-RPSLCYCGKLVKSPN-IDQLASHSLIFQNAFAQQ 81

Db 113 svctpsrsalvgrcirtqvygeeriflptttgiplyevtiaeamkagqyitgmvgk- 171

Qy 82 AVCAPSRVSLTGRPPDTRLTY-DFNSY--WRVHA-GNFS-TIPOYFKENGVTMTSGVKV 136

Db 172 whlginenssdga 185

Qy 137 FHPGISSNHTDDSP 150

RESULT 6

ENTRY JQ1299 #type complete

TITLE N-acetyl galactosamine-6-sulfatase (EC 3.1.6.4) precursor - human

ALTERNATE\_NAMES N-acetyl galactosamine-6-sulfate sulfatase

Result	No.	Score	Query Match	Length	DB	ID	Description	Pred. No.
1	4069	100.0	550	4	IDS	HUMAN	IDURONATE 2-SULFATASE	0.00e+00
2	3129	76.9	563	4	IDS	MOUSE	IDURONATE 2-SULFATASE	0.00e+00
3	282	6.9	497	8	YIDJ	ECOLI	HYPOTHETICAL 57.3 KD	4.30e-32
4	213	5.2	583	7	STS	HUMAN	STERYL-SULFATASE PREC	6.94e-19
5	211	5.2	522	3	GAGS	HUMAN	N-ACETYLGLACTOSAMINE	1.62e-18
6	195	4.8	507	1	ARSA	HUMAN	ARYLSULFATASE A PRECU	1.29e-15
7	190	4.7	533	1	ARSB	HUMAN	ARYLSULFATASE B PRECU	1.01e-14
8	184	4.5	551	1	ARS	HEMPU	ARYLSULFATASE PRECURS	1.15e-13
9	172	4.2	535	1	ARSB	FELICA	ARYLSULFATASE B PRECU	1.39e-11
10	153	3.8	551	1	ASLA	ECOLI	ARYLSULFATASE (EC 5.1	2.10e-08

RESULT	1
ID	IDS HUMAN STANDARD; PRP; 550 AA.
AC	P22304;
DT	01-AUG-1991 (REL. 19, CREATED)
DDT	01-AUG-1991 (REL. 19, LAST SEQUENCE UPDATE)
DT	01-JUN-1994 (REL. 29, LAST ANNOTATION UPDATE)
DE	IDURONATE 2-SULFATASE PRECURSOR (EC 3.1.6.13).
GN	IDS.
OS	HOMO SAPIENS (HUMAN).
OC	EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; MAMMALIA;
OC	EUTHERIA; PRIMATES.
[1]	
RN	SEQUENCE FROM N.A., AND SEQUENCE OF 34-58 AND 456-473.
RC	TISSUE-ENDOTHELIAL CELLS;
RM	91046030
RA	WILSON P.J., MORRIS C.P., ANSON D.S., OCCHIODORO T., BIELICKI J.,
RA	CLEMENTS P.R., HOPWOOD J.J.;
RL	PROC. NATL. ACAD. SCI. U.S.A. 87:8531-8535 (1990).
[2]	
RN	VARIANTS MPS-II ARG-135 AND GLY-422.
RP	93265059
RM	BUNGE S., STEGLICH C., BECK M., ROSENKRANZ W., SCHWINGER E.,
RA	HOPWOOD J.J., GAL A.;
RL	HUM. MOL. GENET. 1:335-339(1992).
RN	[3]

D	b		301	fasyvldtqvrllsealdqlanetlatsdhwalgewakysnfvdathvp <li>  360      </li>
Q	y		301	fasyvldtqvrllsalddlldlanstlaftsdhgawlgewakysnfdvathvp <li>  360      </li>
D	b		361	fypgvtaslpaeagklfpyldpfdeasmepgrqsmdlvelfptlaglqvpp  420 
Q	y		361	fypgvtaslpaeagklfpyldpdfdsasmlpepqrsmdlvelfsptlaglqvpp  420 
D	b		421	rcpwpsfhwclcregnlkkhfrfldespylpgnpreliavsyprpsdiqwnsdkp  480 
Q	y		421	rcpvpsfhwlcregnlkkhfrerdeeelpylcnpnreliaysqprrsdipqwnsdpk  480 
D	b		481	sikdikmgysirtidyrvtyvwgfnpdesflantfsdhagefyfvdscplqdhnmynds <i>q</i> 540 
Q	y		481	sikdikmgysirtidrytwvgfnpdeflanfSDIHAGELYFVDSDP.LQDHNMYNDS <i>Q</i> 540 
D	b		541	ggdlfqllmp  550 
Q	y		541	GCDLFQLLMP  550 
<hr/>				
RESULT			2	
ID	IDS	MUSE	STANDARD;	PRT; 563 AA.
AC	Q08890;	AC		
DT	01-OCT-1994 (REL. 30, CREATED)	DT		
DT	01-OCT-1994 (REL. 30, LAST SEQUENCE UPDATE)	DT		
DT	01-FEB-1995 (REL. 31, LAST ANNOTATION UPDATE)	DT		
DE	IDURONATE 2-SULFATASE PRECURSOR (EC 3.1.6.13).	DE		
GN	IDS.	GN		
OS	MUS MUSCULUS (MOUSE) .	OS		
OC	EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; TETRAPODA; MAMMALIA;	OC		
CC	EUTHERIA; RODENTIA.	CC		
RN	[1]	RN		
RC	SEQUENCE FROM N.A.	RC		
RP	TISSUE=THYMUS;	RP		
RM	93315172	RM		
RA	DANIELE A.; FAUST C.J., HERMAN G.E., DI NATALE P., BALLABIO A.;	RA		
RL	GENOMICS 16:755-757(1993).	RL		
CC	-!- FUNCTION: REQUIRED FOR THE LYSOSOMAL DEGRADATION OF HEPARAN	CC		
CC	SULFATE AND DERMATAN SULFATE.	CC		
CC	-!- CATALYTIC ACTIVITY: HYDROLYSIS OF THE 2-SULFATE GROUPS OF THE	CC		
CC	L-IDURONATE 2-SULFATE UNITS OF DERMATAN SULFATE, HEPARAN SULFATE	CC		
CC	AND HEPARIN.	CC		
CC	-!- SUBCELLULAR LOCATION: LYSOSOMAL.	CC		
CC	-!- SIMILARITY: TO OTHER SULFATASES.	CC		
DR	EMBL; L07921; MMDS.	DR		
DR	PIR; A47153; A47153.	DR		
KW	HYDROLASE; GLYCOPROTEIN; LYSSOME; ZYMOGEN; SIGNAL.	KW		
FT	SIGNAL ?	FT		
FT	PROPEP ?	FT		
FT	CHAIN 60 563	FT		
FT	CARBOHYD 130 130	FT		
FT	CARBOHYD 159 159	FT		
FT	CARBOHYD 261 261	FT		
FT	CARBOHYD 295 295	FT		
FT	CARBOHYD 526 526	FT		
FT	CARBOHYD 550 550	FT		
SC	SEQUENCE 563 AA; 63437 MW; 1700395 CN;	SC		
<hr/>				
DB 4; Score			3129; Match	79.0%; QryMatch 76.9%; Pred. No. 0.00e+00;
Matches			433; Conservative	58; Mismatches 53; Indels 4; Gaps 4;
<hr/>				
D	b		19	araiwrqlsfellglcfcalesaaagsnatdnillivddlrslpcygcklvrsnp  78 :: : :::::   :   :   :    :

DB 8;	Score	282;	Match	28.5%;	QryMatch	6.9%;	Pred. No.	4.30e-32;	
Matches	94;	Conservative	83;	Mismatches	125;	Indels	28;	Gaps	21;
Db	5	nflvmt	dtgatnmvgcysgkplntqnidslaaegirfnsaytcspsvparaglfctgty	64					
Qy	38	NULLIIVDDRISFL-GCYGDKLVSPFNIDQLASHLSLFTQNAFAQAQAVCAPSRVSFLTCRR	96						
Db	65	anqsqpw-t-nnv--apqknistmrgtyfkdayhtcyigk-whldghdyfgtgcceppewd-	119						
Qy	97	POTTRLYDFNSYMRVHAGNFSTIPQYFKENGYSVTMSGVKVFH-PGISNHTDDSPYSWSF	155						
Db	120	adywfdganylse-lt--ekeisilwrnglnsvwdlqanhidetfwahrlnravodflqqp	177						
Qy	156	PYPHPSSKEYNTKCRGPDGELHANLLCPCVDLDV-P-ECTLP-DKQSTEQAIQLLEKM	212						
Db	178	aradepflmvsvdydephhftcpveyleyka-d-fyyelgeakaddl---ankp--e-hh	229						
Qy	213	KYSASPFLLAVGKHPRPIPRYPKFKQLYLENIITLAPDEVPDGLPVPAVPMMDIRO	272						
Db	230	rlwaqamp-s-pvgddgl-yvhhpl---yfcaendfvddqgrvinaltpeqrentwvi-yt	282						
Qy	273	REDVQALNISVPYIGYPVDFQKRIQSTFASVYLDITQVGRLLSALDDQLANSTIATF	332						
Db	283	schgemmgahkkliskgaamydditriplii	312						
Qy	333	SHGWALGEGHEWAKYSN-FDVATHVPLIF	361						

	4	RESULT	
ID	STS HUMAN	STANDARD;	PRT; 583 AA.
AC	P08842;		
DT	01-NOV-1988 (REL. 09, CREATED)		
DT	01-APR-1990 (REL. 14, LAST SEQUENCE UPDATE)		
DT	01-JUN-1994 (REL. 29, LAST ANNOTATION UPDATE)		
DE	STERYL-SULFATASE PRECURSOR (EC 3.1.6.2) (STERIOD SULFATASE) (STERYL-		
DE	SULATE SULFOHYDROLASE) (ARYLSULFATASE C).		
GN	STS.		
OS	HOMO SAPIENS (HUMAN).		
OC	EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; TETRAPODA; MAMMALIA;		
OC	EUTHERIA; PRIMATES.		
[1]	RN		
RP	SEQUENCE FROM N.A.		
RM	89340479		
RA	STEIN C., HILLE A., SEIDEL J., RIJNBOUT S., WAHEED A., SCHMIDT B.,		
RA	GEUZE H., VON FIGURA K.;		
JL	J. BIOL. CHEM. 264:13865-13872 (1989).		
[2]	RN		
RP	SEQUENCE FROM N.A.		
RM	8187642		
RA	YEN P.H., ALLEN E., MARSH B., MOHANDAS T., WANG N., TAGGART R.T.,		
RA	SHAPIRO L.J.;		
RL	CELL 49:443-454 (1987).		
[3]	RN		
RP	SEQUENCE OF 22-45.		
RC	TISSUE=LIVER,		
RM	89352671		
RA	KAWANO J.-I., KOTANI T., OHTAKI S., MINAMINO N., MATSUI H., OINUMA T.,		
RA	AIKAWA E.;		
RL	BIOCHIM. BIOPHYS. ACTA 997:199-205 (1989).		
[4]	RN		
RP	VARIANTS XII LEU-341; ARG-372 AND TYR-446.		
RM	92170784		
RA	BASLER E., GROMPE M., PARENTI G., YATES J., BALLABIO A.;		